

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 04:44:06 ; Search time 20 Seconds  
(without alignments)  
647.031 Million cell updates/sec

Title: US-10-089-014-1

Perfect score: 1646

Sequence: 1 MSDKSAKIFVAGHRLGVSA.....SLRDGLSQTYDWLKNVCR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	943.5	57.3	321	1 FCL_ECOLI	P32055 escherichia
2	917.5	55.7	314	1 Y4AF_RHISN	P55353 rhizobium s
3	510	31.0	312	1 NOLK_AZOCA	P33217 azorhizobium
4	464.5	28.2	321	1 FCL_MOUSE	P23591 m gdp-fucos
5	463.5	28.2	321	1 FCL_HUMAN	Q13630 homo sapien
6	250	15.2	305	1 GALE_METJA	Q57664 methanococc
7	226	13.7	334	1 CAPT_STAAT	P39858 staphylococ
8	219.5	13.3	334	1 GALE_BACHD	Q9kdv3 bacillus ha
9	214	13.0	361	1 RFBB_SALTY	P26391 salmonella
10	211	12.8	329	1 GALE_STRLI	P13226 streptomyce
11	211	12.8	361	1 RFBB_SHIFL	P37777 shigella fl
12	209	12.7	361	1 RBBI_ECOLI	P37759 escherichia
13	208	12.6	338	1 RFPG_HAEIN	P44914 haemophilus
14	206	12.5	341	1 ACBB_ACTS5	Q9zae8 actinoplane
15	204	12.4	361	1 RBBI_ECOLI	P55293 escherichia
16	201	12.2	328	1 STRE_STRGR	P29782 streptomyce
17	193	11.7	341	1 RFBB_RHIMA	Q9s642 neisseria m
18	193	11.7	350	1 RFBB_RHISN	P55462 rhizobium s
19	192	11.7	339	1 GALE_BACSU	P55180 bacillus su
20	190.5	11.6	689	1 GALT_PACTA	P40801 pachysolen
21	190	11.5	315	1 SPSU_BACSU	P39630 bacillus eu
22	190	11.5	699	1 GALT_YEAST	P04397 saccharomyc
23	188.5	11.5	336	1 GALE_YEREN	Q57301 versinia en
24	184	11.2	331	1 GALE_LACCA	Q84903 lactobacill
25	183	11.1	355	1 RFBB_NEIMB	P55294 neisseria m
26	182	11.1	351	1 RFBB_XANCP	P55295 xanthomonas
27	180	10.9	350	1 GAEI_PEA	Q43070 pisum sativ
28	179	10.9	341	1 LPSL_RHIME	O54067 rhizobium m
29	179	10.9	355	1 RFPG_ECOLI	P27830 escherichia
30	178	10.8	346	1 RFBB_NBIGO	P37761 neisseria g
31	177	10.8	329	1 GALE_CORGL	Q45291 corynebacte
32	173	10.5	348	1 EXOB_AZOBR	Q59083 azospirillu
33	172	10.4	350	1 GAE2_CYATE	O65781 cyanomopsis t

## RESULT 1

FCL\_ECOLI  
ID FCL\_ECOLI STANDARD; PRT; 321 AA.  
AC P32055; P76382;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE GDP-fucose synthetase [includes: GDP-mannose-4-keto-6-D epimerase  
(EC 5.1.3.-); GDP-4-keto-6-L-galactose reductase (EC 1.-.-.-)].  
GN FCL OR WCAG OR B2052.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RC STRAIN=K12;  
RX MEDLINE=95115532; PubMed=7815923;  
RA Aoyama K., Haase A.M., Reeves F.R.;  
RT "Evidence for effect of random genetic drift on G+C content after  
lateral transfer of fucose pathway genes to Escherichia coli K-12.";  
RL Mol. Biol. Evol. 11:829-838(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=96326333; PubMed=8759852;  
RA Stevenson G., Andrianopoulos K., Hobbs M., Reeves P.R.;  
RT "Organization of the Escherichia coli K-12 gene cluster responsible  
for production of the extracellular polysaccharide colanic acid.";  
RL J. Bacteriol. 178:4885-4893(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
Saito N., Sampei G., Seki Y., Sivasubram S., Tagami J.,  
Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 40.1-50.0 min region on the linkage map.";  
RL DNA Res. 3:379-392(1996).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=98132401; PubMed=9473059;

## ALIGNMENTS

34	171	10.4	713	1	GAIX_SCHPO
35	170.5	10.4	326	1	YAS5_METJA
36	170	10.3	328	1	EXOB_RHIME
37	170	10.3	328	1	GALE_CORDI
38	168.5	10.2	338	1	GALE_MYCPN
39	166	10.1	334	1	YCL2_ECOLI
40	165.5	10.1	337	1	GALE_ERWAM
41	165	10.0	327	1	EXOB_RHILT
42	165	10.0	351	1	GAEL_ARATH
43	164.5	10.0	338	1	GALE_YERPE
44	162.5	9.9	348	1	GALE_HUMAN
45	161	9.8	350	1	GAEL_ARATH

Q9hd3 schizosacch  
Q58455 methanococc  
P26503 rhizobium m  
P33119 corynebacte  
P75517 mycoplasma  
Q04871 escherichia  
P35673 erwinia amy  
Q59745 rhizobium l  
Q42605 arabidopsis  
Q9f7d4 versinia pe  
Q14376 homo sapien  
Q9sn58 arabidopsis



DE Nodulation protein nolK.  
GN Azorhizobium caulinodans.  
OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Hypomicrobium group; Azorhizobium.  
OX NCBI\_TaxID=7;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ORS571;  
RC MEDLINE=93113009; PubMed=1472710;  
RA Goethals K., Mergaert P., Gao M., Geelen D., van Montagu M.,  
Hollsters M.;  
RT "Identification of a new inducible nodulation gene in Azorhizobium  
caulinodans.";  
RL Mol. Plant Microbe Interact. 5:405-411(1992).  
CC -|- FUNCTION: PROBABLE NUCLEOTIDE SUGAR EPIMERASE/DEHYDROGENASE.  
CC -|- COFACTOR: NAD OR NADP.  
CC -|- INDUCTION: BY FLAVANONE NARINGENIN  
CC -|- SIMILARITY: BELONGS TO THE FUCOSE SYNTHETASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S51942; AAB24744.1; ALT\_INIT.  
DR HSP; P32055; IBSV.  
KW Nodulation; NAD.  
FT NP BIND 6 37 NAD/NADP (POTENTIAL).  
SQ SEQUENCE 312 AA; 34300 MW; BC2FCB01C87C097C CRC64;  
  
Query Match 31.0%; Score 510; DB 1; Length 312;  
Best Local Similarity 35.2%; Pred. No. 6.8e-16;  
Matches 107; Conservative 59; Mismatches 136; Indels 2; Gaps 2;  
  
QY 4 KSAKIFVAGHRLVGSALVRLKQEQFTNLVLKTHAELDLTROADVSEFFSQKPVVIL 63  
DB 3 KKKLLITGGWVGRNLIAACASRGW-EIIATPSVDLRLNAEVEQVIRRLQDPVVH 61  
QY 64 AAKVGGIHNNTYPADFIGVNLQIQTNVIHSAHYEGVKLLFLGSSCIYKFPAPQIPE 123  
DB 62 AAGVGGIHNNTYPADFIGVNLQIQTNVIHSAHYEGVKLLFLGSSCIYKFPACIPELKE 121  
QY 124 SALLTASLEPTNEMWYAIKTAGIKTCQAVRIQHGWDALSGMPTNLGPNNDHPENSHVL 183  
DB 122 CDILRGPFEVNEGALAKTVGLKICEYIDKLPNFNFKTLIACNLYGVGDNFDRSHLL 181  
QY 184 PALMRFPHEAKVNGAEVWVGSGPLREPLHVDLADACVFLLDYRSLG-EHVNIGSGQ 242  
DB 182 PAIEIKHKASQGSSEVSINGDTARREPMFAYDFAKIIKALEVELIPSSMNVGVK 241  
QY 243 EVTIRELAELVKEVVEGKLGWCDTKPDGTPKRLMDSSKASLGWTPKVLSDGLQSQT 302  
DB 242 DLSVLEYISLVARVIGWGSFVYDLNRPVGMRSKLMIDITHLTALGWVPERSEGGIRSTY 301  
QY 303 DWYL 306  
DB 302 QYVI 305  
  
RESULT 4  
FCL\_MOUSE  
ID\_FCL\_MOUSE STANDARD; PRT; 321 AA.  
AC P23591;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE GDP-fucose synthetase (FX protein) (Red cell NADP(H)-binding protein)  
DE (Transplantation antigen P35B) (Tum-P35B antigen) [includes: GDP-  
DE mannose-4-keto-6-D epimerase (EC 5.1.3.-); GDP-4-keto-6-L-galactose

reductase (EC 1.-.-.-)].  
GN TSTA3 OR TSTA35B OR P35B.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT ASN-139.  
RX STRAIN=DBA/2; TISSUE=Maest cells;  
RC MEDLINE=90214611; PubMed=2108859;  
RA Szikora J.-P., van Pel A., Brichard V., Andre M., van Baren N.,  
Henry P., de Plaen E., Boon T.;  
RT "Structure of the gene of tum- transplantation antigen P35B: presence  
RT of a point mutation in the antigenic allele.";  
RL EMBO J. 9:1041-1050(1990).  
RN [2]  
RP REVISIONS.  
RA De Plaen E.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: TWO STEP NADP-DEPENDENT CONVERSION OF GDP-4-DEHYDRO-6-  
CC DEOXY-D-MANNOSE TO GDP-FUCOSE, INVOLVING AN EPIMERASE AND A  
CC REDUCTASE REACTION (BY SIMILARITY).  
CC -|- PATHWAY: Conversion of GDP-mannose to GDP-fucose; second step.  
CC -|- PATHWAY: Conversion of GDP-mannose to GDP-fucose; third (last)  
CC step.  
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -|- MISCELLANEOUS: MUTAGEN TREATMENT OF P815 TUMOR CELLS PRODUCES TUM-  
CC VARIANTS THAT ELICIT A CYTOLYTIC T LYMPHOCYTE RESPONSE (CTL). THE  
CC ANTIGENIC ALLELE DIFFERS FROM THE NORMAL ALLELE BY A SINGLE  
CC MUTATION IN POSITION 139.  
CC -|- SIMILARITY: BELONGS TO THE FUCOSE SYNTHETASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X53620; CAB94217.1; ALT\_SEQ.  
DR EMBL; X53621; CAB94217.1; JOINED.  
DR EMBL; X53622; CAB94217.1; JOINED.  
DR EMBL; X53623; CAB94217.1; JOINED.  
DR EMBL; X53624; CAB94217.1; JOINED.  
DR EMBL; X53625; CAB94217.1; JOINED.  
DR EMBL; X53626; CAB94217.1; JOINED.  
DR EMBL; X53627; CAB94217.1; JOINED.  
DR EMBL; X53628; CAB94217.1; JOINED.  
DR EMBL; M30127; AAA39673.2;  
DR EMBL; M30128; AAA39674.1; ALT\_FRAME.  
DR PIR; S12516; S12516.  
DR MGD; MGI:98857; Tsta35b.  
KW Isomerase; Oxidoreductase; NADP; Multifunctional enzyme;  
KW Tumor antigen; Antigen.  
FT NP BIND 9 39 NADP (POTENTIAL).  
FT VARIANT 139 139 S -> N (IN TUM- ALLELE).  
SQ SEQUENCE 321 AA; 35877 MW; 358D86D68F173531 CRC64;  
  
Query Match 28.2%; Score 464.5; DB 1; Length 321;  
Best Local Similarity 36.0%; Pred. No. 5e-32;  
Matches 113; Conservative 57; Mismatches 129; Indels 15; Gaps 6;  
  
QY 5 SAKIFVAGHRLVGSALVRLKQEQF- - -TNVLKTHAELDLTROADVSEFFSQKPVY 60  
DB 7 SMRLVTGSGLVGRAL-OKVADGAGLPGEEWVFVSSKDDADLTDAQOALQKQVPTH 65  
QY 61 VILAAKVGGIHNNTYPADFIGVNLQIQTNVIHSAHYEGVKLLFLGSSCIYKFPAPQ 120  
DB 66 VIHLAAWVGGLFENIKYNLDLFWKKNVHNDVNLHSAFEGVARKVSVCLSTCIIPDKTTP 125  
QY 121 IPESALLTASLEPTNEMWYAIKTAGIKTCQAVRIQHGWDALSGMPTNLGPNNDHPEN 180  
DB 122 IPESALLTASLEPTNEMWYAIKTAGIKTCQAVRIQHGWDALSGMPTNLGPNNDHPEN 180

Db 126 IDMTIHNGPPHNSFGYSYAKRMIDVQNRAYFQOHCCTFTAVIPTNVFGPDNFNIEDG 185  
QY 181 HVLPAALMRREHAKVNGAEVWVWGTGSPLEFLHVDDLDADACVFLLDYSGLEHV--NI 238  
Db 186 HVLPLGLHKVHLAK-SSDSALTWGCKPRQIYSIDLARLFVWLREYSEVEPIILSV 244  
QY 239 GSGQEVITRELAELVKEVVGFEGLKWDCTKPGTGRKMDSSKSLAS-----LGTWPKVSL 294  
Db 245 GEEDEVSIEKAAEAHVVEAMDENGVEVTFDTSKDGQYKKTASNGKLSYLPDFRFTF---F 301  
QY 295 RDGLSQTVDWYLN 308  
Db 302 KOAVKETCAWFTDN 315

RESULT 5  
FCL\_HUMAN  
ID FCL\_HUMAN STANDARD; PRT; 321 AA.  
AC Q13630; Q9UDG7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE GDP-fucose synthetase (FX protein) (Red cell NADP(H)-binding protein)  
DE [includes: GDP-mannose-4-keto-6-D epimerase (EC 5.1.3.-); GDP-4-keto-6-L-galactose reductase (EC 1.-.-.-)]  
GN TSTA3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND SUBUNIT.  
RC TISSUE=Placenta;  
RX MEDLINE=97066899; PubMed=8910301; Benatti U., De Flora A.;  
RA Tonetti M., Sturla L., Basso A., Benatti U., De Flora A.;  
RT "Synthesis of GDP-L-fucose by the human FX protein.";  
RL J. Biol. Chem. 271:27274-27279(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX Strausberg R.;  
RA Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP PARTIAL SEQUENCE.  
RT TISSUE=Erythrocyte;  
RX MEDLINE=95102076; PubMed=7803801;  
RA Camardella L., Carratore V., Ciardiello A., Damonte G., Benatti U.,  
De Flora A.;  
RT "Primary structure of human erythrocyte nicotinamide adenine dinucleotide phosphate (NADP(H))-binding protein FX: identification with the mouse tum-transplantation antigen P35B.";  
RL Blood 85:264-267(1995).  
CC -!- FUNCTION: TWO STEP NADP-DEPENDENT CONVERSION OF GDP-4-DEHYDRO-6-DEOXY-D-MANNOSE TO GDP-FUCULOSE, INVOLVING AN EPIMERASE AND A REDUCTASE REACTION.  
CC -!- PATHWAY: Conversion of GDP-mannose to GDP-fucose; second step.  
CC -!- PATHWAY: Conversion of GDP-mannose to GDP-fucose; third (last) step.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- SIMILARITY: BELONGS TO THE FUCULOSE SYNTHETASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U58766; AAC50786.1; -;  
CC EMBL; BC001941; AA01941.1; -;  
CC Genew; HGNC:12390; TSTA3.  
CC MIM; 137020; -;

KW Isomerase; Oxidoreductase; NADP; Multifunctional enzyme.  
FT NP BIND 5 39 NADP (POTENTIAL)  
FT CONFLICT 5 Q -> E (IN REF. 3).  
SQ SEQUENCE 321 AA; 35892 MW; 94BB1PF61658007C CRC64;  
Query Match 28.2%; Score 463.5; DB 1; Length 321;  
Best Local Similarity 35.4%; Pred. No. 6.1e-32;  
Matches 111; Conservative 64; Mismatches 124; Indels 15; Gaps 6;  
QY 5 SAKIFVAGHGLVGSIAVRKLEQCGF---TNVLKTHAELDLTROADVSEFFSOEKPVY 60  
Db 7 SMRIIVTGSGLVGNKAI-OKVVADGAGLPGEDWVVFSSKADADLTDTATQTRALFEKVQPTH 65  
QY 61 VILAAAKVGGIHANNTPADFIGVNLQIQTNVIHSAYEHGVKGLFLGLSSCIYKFAPOP 120  
Db 66 VIHLAAMVGLFRNIKYNLDFWRKVMNDNVLSHSAFVGARKVVSCLCIFPDKTYP 125  
QY 121 IPESALLTASLEPTNWEYAIKIQTCQAYRIQHGWDATSCMTNLYGPDNHPENS 180  
Db 126 IDETMHNHNGPPHNSFGYSYAKRMIDVQNRAYFQOYCGCTFTAVIPTNVFGPDNFNIEDG 185  
QY 181 HVLPAALMRREHAKVNGAEVWVWGTGSPLEFLHVDDLDADACVFLLDYSGLEHV--NI 238  
Db 186 HVLPLGLHKVHLAKSSGS-ALTVMGTGMPRRQFIYSLDAQLFVWLREYNEVEPIILSV 244  
QY 239 GSGQEVITRELAELVKEVVGFEGLKWDCTKPGTGRKMDSSKSLAS-----LGTWPKVSL 294  
Db 245 GEEDEVSIEKAAEAHVVEAMDENGVEVTFDTSKDGQYKKTASNGKLSYLPDFRFTF---F 301  
QY 295 RDGLSQTVDWYLN 308  
Db 302 KOAVKETCAWFTDN 315

RESULT 6  
GALE\_METJA  
ID GALE\_METJA STANDARD; PRT; 305 AA.  
AC Q57664;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-glucose 4-epimerase).  
GN MJ0211.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=868087;  
RA Sutton G.G., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -!- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.  
CC -!- COFACTOR: NAD.  
CC -!- PATHWAY: Galactose metabolism; third step.  
CC -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U58766; AAC50786.1; -;  
CC EMBL; BC001941; AA01941.1; -;  
CC Genew; HGNC:12390; TSTA3.  
CC MIM; 137020; -;



CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
---  
EMBL; U0927; AAA64648.1; -.  
HSSP; P27830; 1BXK.  
InterPro; IPR001509; Epimerase\_Dh.  
PFam; PF01370; Epimerase; 1.  
Lyase; NAD.  
KW  
SQ SEQUENCE 334 AA; 37931 MW; 2FDFBF8D4D7796CF CRC64;

Query Match	13.7%	Score 226;	DB 1;	Length 334;
Best Local Similarity	21.4%;	Pred. No. 8.1e-12;		
Matches 76;	Conservative	71;	Mismatches 128;	Indels 80;
Gaps 11;				

2	KIDILIGAGFISGHEKAKLQKHVYVGDSDINDIITSVSKUEKRLKSIGNETIFPNKVAL	81	
QY	44	TROADYESFFSQBPVYVILAAKVGSGIHANNTYPADFIGVNIQIQITNVIHSAYEHGVKK	103
Db	62	ENYDDLKSVFVDSQPEVVVNLAAQ-AGVRYSIENPRTYIDSNIVGFWMNILLECRSHFNQ	120
QY	104	LLPLGSSCTYKPAPOP-----IPESALLTASLEPTNEWATAKTAGIKTCQAYRIQHG	157
Db	121	LIYASSSSVYGANTSKPFSTDNDIDHPLSLYAAATKKSNELMA-----HT	164
QY	158	WDAISGMP-----NLYGPNDFHPENSHVLPALMRRPHEAKVNGAEVVVWGTSPLR	211
Db	165	YSHLYNLPTGLRFFTVYGPWG--RPDMA-----LFKFTKALVND-QAIDVYHNGNMR	215
QY	212	EFHVLVDLADACVFLDLR-----YSGLEHVNISGSGQVTTIRELAELV	253
Db	216	DFTYVDIIVEAISRLYKPKASPKNKESGADPDGSSSVAPKYVYNI GNNSPVRLMEFVEAI	275
QY	254	KEVVGEGKLGWDCThPDGTPRKLMDSKL-ASLGWTPKVSRLRDLGSQTVDWYLK	307
Db	276	ENKLGKEARKNYMDLQGVDPETPYANVDLFRIDIEKPTTIQDGVNKKFVDWYLE	330

GALE	BACHD	STANDARD;	PRI;	334	AA.
Q9KDV3;					
16-OCT-2001	(Rel. 40, Created)				
DT	AC				
DT	DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-				
DE	galactose 4-epimerase).				
GN	GALE OR BH1108.				
OS	Bacillus halodurans.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=86665;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C-125 / JCM 9153;				
RX	MEDLINE=20512582; PubMed=11058132;				
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Maui N.,				
RA	Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,				
RA	Horikoshi K.;				
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus				
RT	halodurans and genomic sequence comparison with Bacillus subtilis.";				
RL	Nucleic Acids Res. 28:4317-4331(2000).				
CC	-1- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.				
CC	-1- COFACTOR: NAD.				
CC	-1- PATHWAY: Galactose metabolism; third step.				
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.				

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities is allowed on request.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AP001510; BAB04827.1; -.  
DR HSP; P09147; IKVS.  
DR InterPro; IPR001509; Epimerase Dh.  
DR InterPro; IPR000205; NAD binding.  
DR Pfam; PF01370; Epimerase; 1.  
DR TIGRFAMs; TIGR01179; Gale; 1.  
KW Isomerase; NAD; Galactose metabolism; Complete proteome.  
FT NP\_BIND 2 33 NAD (POTENTIAL).  
SQ SEQUENCE 334 AA; 37077 MW; 2CDEFA3DE7336EC14 CRC64;

Query Match 13.3%; Score 219.5; DB 1; Length 334;  
Best Local Similarity 27.0%; Pred. No. 2.9e-11;  
Matches 92; Conservative 57; Mismatches 125; Indels 67; Gaps 18;

QY 8 IFVAGHRLVGSIAVRKLOQGFNLVL-----KTHAEL-----DLTRQADVESFFS 54  
DB 3 ILVTGGAGYIGSTVLFLLBQGEQVIVLDNLQKHAGALSVDVTFYHGDIRDQDLDTIFT 62  
QY 55 QEKPVVILLAAK--VGGIHANNYPADFGVNLQITNVHSAIEHGVKLLFLGSSCI 112  
DB 63 THSIDTVIHPAANSLVG---ESVKQPIEYENNVGTHLLKKQLEHDKVKIVFSTAAT 119  
QY 113 YKFAFOPIPESSALLTASLEPTNEWYAIAKIAGIKT---CQ-AYRIQH---GWDALSGM 164  
DB 120 YGEPVQIPIQES---DPTIPTNP-YGETKLAIEKMFHWCQEAIGLYVCLRYFNAGAD 174  
QY 165 PTNLYGNDFHNPENSHVLPALMRPHEAKVNGAEVVVWG-----TGSPLREFLHVD 218  
DB 175 PNGRIGED---HSPESHLPVL---QVALGQRERVAIFGDYQTEDGSCIRDYIHMD 227  
QY 219 LADACVFLLDYRYSGLHV-----NIGSGQEVIRIELAELVKEVWG---FEGKLGWDC 267  
DB 228 LANA-----HYLACSHLRKDGSGFNLGNGKGFVSEVIEVCROVTHGPIPAEI---A 278  
QY 268 TKPDGTFRKLMDSKSLAS--LGWTPKV-SLRDGLSOTYDWY 305  
DB 279 PRSGDPASLIASSERAQITLGWEKYPKSLETWHEANWH 319

RESULT 9  
RFBF\_SALTY STANDARD; PRT; 361 AA.  
ID RFBF\_SALTY  
AC P26351;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE dTDP-glucose 4,6-dehydratase (EC 4.2.1.46).  
GN RFBF OR STM2097.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2;  
RX MEDLINE=91260454; PubMed=1710759;  
RA Jiang X.-M., Neal B., Santiago F., Lee S.-J., Romana L.K., Reeves P.R.;  
RT "Structure and sequence of the rfb (O antigen) gene cluster of  
RT Salmonella serovar typhimurium (strain Lt2).";  
RL Mol. Microbiol. 5:695-713(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

RT LT2.";  
RL Nature 413:852-856(2001).  
CC -1- CATALYTIC ACTIVITY: dTDP-glucose = dTDP-4-dehydro-6-deoxy-D-  
CC glucose + H(2)O.  
CC -1- COFACTOR: NAD.  
CC -1- PATHWAY: DTPD-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN  
CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. DTPD-GLUCOSE  
CC DEHYDRATASE SUBFAMILY.  
CC -----

This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X56793; CAA40115.1; -.  
DR EMBL; AE008792; AAL21001.1; -.  
DR PIR; S15299; S15299.  
DR HSP; P27830; IBXK.  
DR StyGene; SGI0345; rfbB.  
DR InterPro; IPR001509; Epimerase Dh.  
DR Pfam; PF01370; Epimerase; 1.  
DR TIGRFAMs; TIGR01181; dTDP gluc dehyd; 1.  
KW Lipopolysaccharide biosynthesis; Lyase; NAD; Complete proteome.  
FT NP\_BIND 7 13 NAD (POTENTIAL).  
SQ SEQUENCE 361 AA; 40718 MW; 3A574B4D917BEC57 CRC64;

Query Match 13.0%; Score 214; DB 1; Length 361;  
Best Local Similarity 22.7%; Pred. No. 9.3e-11;  
Matches 83; Conservative 58; Mismatches 132; Indels 92; Gaps 13;

QY 7 KIFVAGHRLVGSIAVR---KLQEGFTNLVLKTHA-----ELDLTR 45  
DB 2 KILITGGAGFIGSAVVRHIIKNTQTVVNIIDKLTAGNLSLSDISESNRYNFEHADICD 61  
QY 46 QADVESFFSQKPVVILAAAKVGGIHANNYPADFGVNL-----86  
DB 62 SAEITRIFEQYQPDAMVHLAAE-SHVORSITGPAFIETNIIVGTVALLEVARKYWSALGE 120  
QY 87 -----QITNVHSAIEHGVKLLFLGSSCIYKFAFOPIPESSALLTASLEPT--- 134  
DB 121 DKCNFRFHISTDEVYGLPH-----PDEVNSVTLPFTTTAYA 162  
QY 135 -NEWYAIAGIKTCQAYRIQHGWDALSGMPTNLYGNDFH-PENSHVLPALMRPHE 192  
DB 163 PSSPYASASKASDHLVRAWRRTYGLPTIVTNCNNYGP---YHFPEK--LIPLVILNALE 217  
QY 193 AKVNGAEVVVWGTSPLREFLHVDLADACVFLLDYRYSGLHVNIGSGQE-----VT 245  
DB 218 GK-----PLPIYKGDQIRDLVYEDHARALHVMVTEKAGETVYNIHGHNKKNLDVVFT 272  
QY 246 IRELA-ELVKEVVGPEGLKWDCTKPDCTPRKLMDSKSLA-SLGWTPKVSRLDGLSQTVD 303  
DB 273 ICOLLDEIVPKATSVREQITVYVADRPCHGDRYDAIDAGKISRELGWKPLETFESGIRKIVE 332  
QY 304 WYLNK 308  
DB 333 WYLAN 337

RESULT 10  
GALE\_STRLI STANDARD; PRT; 329 AA.  
ID GALE\_STRLI  
AC P13226;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE UDP-glucose 4-epimerase (EC 5.1.1.3.2) (Galactowaldenase) (UDP-  
DE galactose 4-epimerase).  
GN GALE.



Db 61 DAVAMSRIFAQHPDAMVHLAAE-SHVDRSITGPAAFIETNI-VGTVVLLEARNYWSAL 118  
 QY 98 EHGKLLFLGSSCIYKPAQIPESALITASLEPTNEWYATAKIGTKOAY--RIQ 155  
 Db 119 NDEKKKSRFHHSITDEVYGDLPHPDEANNEALPLFTTAYAPSSPYASAKASDHLV 178  
 QY 156 HGWDAISGMT-----NLYGPNDFH-PENSHVLPALMRPFHEAKVNGAEVVVWGTS 208  
 Db 179 RAMKRTYGLPTIVTNCNNYGP---YHFPEK--LIPVLINALEGKA-----LPIYKGD 228  
 QY 209 PLREFLHVDDADACVFLDRLYSGLEHVNIGSQE-----VTIRELA-ELVKEVVGFE 260  
 Db 229 QIRDWLYVEDHARALYTVVTEGKAGETYNIGHNEKNIDVLTICDLDLDEIVPKEKSYR 288  
 QY 261 KGLGWDCTKDPGTPKRLMSSKLA-SLGWTPKVSRLDGLSOTYDYLKN 308  
 Db 289 EQITYVADRPGRHRRYDAIDAKISRELGMKQPQTFESGIRKTVWYLAN 337

## RESULT 12

RBBI\_ECOLI  
 ID RBBI\_ECOLI STANDARD; PRT; 361 AA.  
 AC P3759; P78082;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE dTDP-glucose 4,6-dehydratase (EC 4.2.1.46).  
 GN RFB OR RMLB OR B2041.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / WGI;  
 RX MEDLINE=94292435; PubMed=7517391;  
 RA Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M.,  
 RA Redmond J.W., Lindquist L., Reeves P.R.;  
 RT "Structure of the O antigen of Escherichia coli K-12 and the sequence  
 of its rfb gene cluster.";  
 RL J. Bacteriol. 176:4144-4156(1994).  
 RN [2]  
 RP REVISIONS TO 123 AND 250.  
 RC STRAIN=K12 / WGI;  
 RA Stevenson G.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251358; PubMed=9097040;  
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,  
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,  
 RA Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,  
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;  
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 40.1-50.0 min region on the linkage map.";  
 RL DNA Res. 3:379-392(1996).  
 CC -!- CATALYTIC ACTIVITY: dTDP-glucose = dTDP-4-dehydro-6-deoxy-D-  
 glucose + H<sub>2</sub>O.  
 CC -!- COFACTOR: NAD.

CC -!- PATHWAY: dTDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN  
 BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. dTDP-GLUCOSE  
 DEHYDRATASE SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; U09876; AAB88398.1; -.  
 DR EMBL; AE000294; AAC75102.1; -.  
 DR EMBL; D90841; BAA15883.1; -.  
 DR EMBL; D90842; BAA15895.1; -.  
 DR HSSP; F27830; 1BXX.  
 DR EcoGene; EG12412; rfbB.  
 DR InterPro; IPR001509; Epimerase\_Dh.  
 DR Pfam; PF01370; Epimerase\_1.  
 DR TIGRFAMs; TIGR01181; dTDP-gluc\_dehyd; 1.  
 KW Lipopolysaccharide biosynthesis; lyase; NAD; Complete proteome.  
 FT NP\_BIND 7 13 NAD (POTENTIAL).  
 FT CONFLICT 123 123 K -> E (IN REF. 1).  
 FT CONFLICT 250 250 G -> V (IN REF. 1).  
 SQ SEQUENCE 361 AA; 40558 MW; C3F64643271C14C7 CRC64;  
 Query Match 12.7%; Score 209; DB 1; Length 361;  
 Best Local Similarity 25.4%; Pred. No. 2.5e-10;  
 Matches 89; Conservative 59; Mismatches 139; Indels 64; Gaps 17;  
 QY 7 KIFVAGHGLVGSATVRKL---QEQFTNLVLKTHA-----ELDLTRADV-- 49  
 Db 2 KILVTGGAGFISAVVRHHIINTQSVNVNPKLTYAGNRESLADVSDSERYVFEHADICD 61  
 QY 50 ----ESFFSQEKPVVYVILAAAKVGIHANNTYPADFIGNVLIQIOTNVHSAHEH----- 99  
 Db 62 APAMARIFAQHPDAMVHLAAE-SHVDRSITGPAAFIETNI-VGTVVLLEARNYWSALD 119  
 QY 100 GVKLLP-----LGSSCIYKPAQIPESALITASLEPTNEWYATAKIGTKOAY--R 153  
 Db 120 SDKKNSFRFHHISTDEVY---GDLPHDPDEVNTELPFTTAYAPSSPYASAKASD 176  
 QY 154 IQHGWDAISGMT-----NLYGPNDFH-PENSHVLPALMRPFHEAKVNGAEVVVWG 206  
 Db 177 LVRAWKRTYGLPTIVTNCNNYGP---YHFPEK--LIPVLINALEGKA-----LPIYK 226  
 QY 207 GSPLEFLHVDDADACVFLDRLYSGLEHVNIGSQE-----VTIRELA-ELVKEVVG 258  
 Db 227 GDQIRDWLYVEDHARALYTVVTEGKAGETYNIGHNEKNIDVLTICDLDLDEIVPKEKS 286  
 QY 259 FEGKLGWDCTKDPGTPKRLMSSKLA-SLGWTPKVSRLDGLSOTYDYLKN 308  
 Db 287 YREQITYVADRPGRHRRYDAIDAKISRELGMKQPQTFESGIRKTVWYLSN 337  
 RESULT 13  
 RFFG\_HAEIN  
 ID RFFG\_HAEIN STANDARD; PRT; 338 AA.  
 AC P44914;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE dTDP-glucose 4,6-dehydratase (EC 4.2.1.46).  
 GN RFFG OR H10873.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;











```

; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; FILE REFERENCE: BASP-NAE-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patentin Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 16
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-09-734-569-16

Query Match 40.8%; Score 671; DB 10; Length 162;
Best Local Similarity 74.8%; Pred. No. 1.1e-58;
Matches 119; Conservative 25; Mismatches 15; Indels 0; Gaps 0;

QY 150 QAYRQHGWDASGMPNLYGNDNPHNSHVLPALMRPFHEAKVNGAEVVMWGTSQP 209
DB 1 QAYRLQYNFDAISGMPNLYGNDNPHNSHVLPALMRPFHEAKVNGAEVVMWGTSQP 60
QY 210 LREFLHVDLADACVFLDRLYSLGHEVNTGSGOEVITIRELAELVKEVVGFEKLGWDCTK 269
DB 61 FREFLHVDLADACVFLDRLYSLGHEVNTGSGOEVITIRELAELVKEVVGFEKLGWDCTK 120
QY 270 PDGTPRKLMDSSKSLASLWTPKVSRLDGLSQTVDWYLN 308
DB 121 PDGTPRKLMDSSKSLASLWTPKVSRLDGLSQTVDWYLN 159

RESULT 5
US-09-318-271-6
; Sequence 6, Application US/09318271A
; Patent No. US20020012979A1
; GENERAL INFORMATION:
; APPLICANT: Berry, Alan
; APPLICANT: Running, Jeffrey A.
; APPLICANT: Sevenson, David K.
; APPLICANT: Burlingame, Richard P.
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"
; FILE REFERENCE: 3161-24
; CURRENT APPLICATION NUMBER: US/09/318,271A
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 60/125,073
; EARLIER FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 60/125,054
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 60/088,549
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-318-271-6

Query Match 28.2%; Score 463.5; DB 10; Length 321;
Best Local Similarity 35.4%; Pred. No. 1e-37;
Matches 111; Conservative 64; Mismatches 124; Indels 15; Gaps 6;

QY 5 SAKIFVAGHGLVGSIAVRKLEQGF---TNLVKTHAELDLTROADVSEFSOEKPVY 60
DB 7 SMRLVTGSGGLVGRKAI-QKVADGAGLCEGDEWVFVSSKADLDTAQTALFEKVPQPH 65
QY 61 VILAAKVGGIHANNTPADFTGVNLQIQTNVHSAHYHGKLLFLGSSCIYKPFAPQP 120
DB 66 VIHLAAWUGLFRNIKNYLDLFWKKNVHMNDNLHSAFEVARGKAVSCISLCIFDKTTP 125
QY 121 IPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDASGMPNLYGNDNPHNS 180

```

```

DB 126 IDETHIHNGPPHNSNFGSYAKRMIDVQNRAYFOQYGCFTTAVITNVFPGPHDNFIEDG 185
QY 181 HVLPALMRPFHEAKVNGAEVVMWGTSQPLREFLHVDLADACVFLDRLYSLGHEV--NI 238
DB 186 HVLPLIHKVHLAKSSGS-ALTVMGTGNPRQFIYSLDLAQLFIWLVREYNEVEPIILSV 244
QY 239 GSGQEVITIRELAELVKEVVGFEKLGWDCTKPDGTPRKLMDSSKSLAS-----LGTWPKVSL 294
DB 245 GSEDEVSISKEAAEAIVEAMDPEHGEVTFDTTKSDGQFKKTASNSKLRTYLPDRFETP---F 301
QY 295 RDGLSQTVDWYLN 308
DB 302 KOAVKETCAWFTDN 315

RESULT 6
US-09-738-626-3862
; Sequence 3862, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 3862
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3862

Query Match 15.1%; Score 249; DB 9; Length 311;
Best Local Similarity 25.3%; Pred. No. 1.8e-16;
Matches 82; Conservative 55; Mismatches 139; Indels 48; Gaps 9;

QY 10 VAGHGLVGSIAVRKLEQGFNTLVK-----THAELDLTROADVE 50
DB 5 VTGAGFTGSHLVLDLLKEGHEVVVDNLSSRGLNLSDAEATGKLTFFVADLL-DVDNF 63
QY 51 SFFSOEKPVYVTLAAAKVGGIHANNTPADFTGVNLQIQTNVHSAHYHGKLLFLGSG- 109
DB 64 EFLGTHKEVTFHLAAQIDVRH-SVVDPLHDAETHILSTIRADAAHQGVKRVFTSSG 122
QY 110 SCYKPKFAPQIPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDASGMPNLY 169
DB 123 GSIYGESEFPVDE---TVPVDHSP-YAASKVSGEYIYNTFRHLYGLDCSHIAPANVY 177
QY 170 GENDNPHNSHVLPALMRPFHEAKVNGAEVVMWGTSQPLREFLHVDLADACVFLDRL 229
DB 178 GROPDPHGEAGVAJFALR-----LLGGLDTPKFGDGNTRDYVYVGVWVRAFYLASGE 231
QY 230 YSLGSHVNIQSQEVITIRELAELVKEVVGFEKLGWDCTKPDGTPRKLMDSSKSLA----- 284
DB 232 IGGGERFNIGTSVETSDRLHLTLVATAAGSK-----DDPEYAPARLGDVPRSAISFK 284
QY 285 ---SLGWTPKVSRLDGLSQTVDWY 305

```



;; PRIOR APPLICATION NUMBER: 60/064809  
;; PRIOR FILING DATE: 1997-11-07  
;; PRIOR APPLICATION NUMBER: 60/065186  
;; PRIOR FILING DATE: 1997-11-12  
;; PRIOR APPLICATION NUMBER: 60/065846  
;; PRIOR FILING DATE: 1997-11-17  
;; PRIOR APPLICATION NUMBER: 60/066364  
;; PRIOR FILING DATE: 1997-11-21  
;; PRIOR APPLICATION NUMBER: 60/066453  
;; PRIOR FILING DATE: 1997-11-24  
;; PRIOR APPLICATION NUMBER: 60/066511  
;; PRIOR FILING DATE: 1997-11-24  
;; PRIOR APPLICATION NUMBER: 60/066770  
;; PRIOR FILING DATE: 1997-11-24  
;; PRIOR APPLICATION NUMBER: 60/069212  
;; PRIOR FILING DATE: 1997-12-11  
;; PRIOR APPLICATION NUMBER: 60/069278  
;; PRIOR FILING DATE: 1997-12-11  
;; PRIOR APPLICATION NUMBER: 60/069334  
;; PRIOR FILING DATE: 1997-12-11  
;; PRIOR APPLICATION NUMBER: 60/069694  
;; PRIOR FILING DATE: 1997-12-16  
;; PRIOR APPLICATION NUMBER: 60/072320  
;; PRIOR FILING DATE: 1998-01-23  
;; PRIOR APPLICATION NUMBER: 60/073612  
;; PRIOR FILING DATE: 1998-02-04  
;; PRIOR APPLICATION NUMBER: 60/074086  
;; PRIOR FILING DATE: 1998-02-09  
;; PRIOR APPLICATION NUMBER: 60/074092  
;; PRIOR FILING DATE: 1998-02-09  
;; PRIOR APPLICATION NUMBER: 60/077791  
;; PRIOR FILING DATE: 1998-03-12  
;; PRIOR APPLICATION NUMBER: 60/078910  
;; PRIOR FILING DATE: 1998-03-20  
;; PRIOR APPLICATION NUMBER: 60/079294  
;; PRIOR FILING DATE: 1998-03-25  
;; PRIOR APPLICATION NUMBER: 60/079663  
;; PRIOR FILING DATE: 1998-02-27  
;; PRIOR APPLICATION NUMBER: 60/079728  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/080165  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/081203  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081229  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081695  
;; PRIOR FILING DATE: 1998-04-14  
;; PRIOR APPLICATION NUMBER: 60/081817  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081818  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082999  
;; PRIOR FILING DATE: 1998-04-24  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083545  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/084600  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084627  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084637  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085149  
;; PRIOR FILING DATE: 1998-05-12  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085579

;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/086414  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/086430  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087106  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/088026  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088730  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088741  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088810  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088858  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089532  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089599  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089907  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089947  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/090349  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090429  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090538  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07

Query Match 13.1%; Score 215; DB 9; Length 420;

Best Local Similarity 25.3%; Pred. No. 6.6e-13; Mismatches 138; Indels 66; Gaps 13;  
Matches 87; Conservative 53;

QY 1 MSDKSAK-IFVAGHRLVGSATVRKLOEQGFTNLVLT-----HAELDLTR 45  
Db 83 LSEKORKRILITGGAGFGVSHLTDKLMMDGHEVTVVDNFFTKRKNVHEWIGHENFELIN 142  
QY 46 QADVESFFSQEKPVPVILAAKVGGIHANTYTPADFIGVNLQ-IQNVIHSAVEHGKK- 103  
Db 143 HDVVEPLYIEVDQIVHLASPAS-----PPNYMNPPIKTKNTIGTLNMLGLAKR 192  
QY 104 ----LLFLGSSCIYKFAQPIPEESALLTASLEP--TNEWYATAKTAGIKTCOAYRIOHG 157  
Db 193 VGARLLLASTSEVYGD--PEVHPQSEDYGHVGNPIGPRACYDEGKRVATMCTAYMKQEG 250  
QY 158 WDAISGMPTNLXGP-----NDNFHPENSHVLPALMRRFHEAKVNGABEVVWGTGSPREF 213  
Db 251 VEVRVARIENTFGPRMHMDG-RVVSNFLQALQ-----GEPLTVYGSQSOTRAF 299  
QY 214 LHVDDLADACVFLLDTRYSGLEHVNIGSGOEVTIRELAELVKEVVGEGKLGW----- 265  
Db 300 QVVSOLVNGLVALMNS-NVSSPVNLGNPEEHTILEPAQLIKNLVGSSEIQTFLSEAQDDP 358  
QY 266 DCTKPDGTPRKLMDSSKLASIGWTPKVSRLDGLSOTDYWLKNV 309  
Db 359 QKRKPDIKKAKLM-----LGWPEVYVPLEEGLNKAHYFRKEL 395

## RESULT 9

US-10-121-049-260

; Sequence 260, Application US/10121049

; Publication No. US2003002239A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C17

; CURRENT APPLICATION NUMBER: US/10/121,049

; CURRENT FILING DATE: 2002-04-12

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 260

; LENGTH: 420

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-121-049-260

Query Match

Best Local Similarity 25.3%; Score 215; DB 9; Length 420;

Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAK-IFVAGHRLVGSALVRKLOEQGFNTLVKLT-----HAELDLTR 45

Db 83 LSEKDRKRILITGGAGFVGSHTDKLMDGHEVTVVDNFTGKRKNVHEWHIGHENFELIN 142

QY 46 QADVESFFSQEPKPVYVILAAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103

Db 143 HDVVEPLYEVDQIYHLASPAS-----PPNYMNPITKLTNTIGTLMGLAKR 192

QY 104 ----LLFLGSSCIYPKFAPOPIPESSALLTASLEP--TNEWYAIAGIKTKQAYRIQHG 157

Db 193 VGARLLLASTSEVYGD--PEVHPQSEDYWGHNPIGPACRYDEGKRVAETMTCYAYMKQSG 250

QY 158 WDAISGMPTNLVGP-----NDNFHPENSHVLPALMRFPHEAKVNGAEVVVWGTGSPLERF 213

Db 251 VEVRVARIFTFGPRMMDNG-RVVSNFILQALQ-----GEPLTVYSGSGQTRAF 299

QY 214 LHVDDLADACVPLLDVRYSGLEHVNIGSGQEVTTIRELAEVLKVEVFGKLGW----- 265

Db 300 QYVSDLVNGLVALMNS-NVSSPVNLGNPEHTILEFAQLIKNLVSGSGSEIQFLSEAQDDP 358

QY 266 DCTKPDGTTPRKLMDSSKSLASLGWTPKVSRLDGLSQTQYDWYLNK 309

Db 359 QKRKPDIIKAKLM-----LGWEPVPLEEGLNKAHYFRKEL 395

## RESULT 10

US-10-123-904-260

; Sequence 260, Application US/10123904

; Publication No. US20030022328A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C54  
; CURRENT APPLICATION NUMBER: US/10/123,904  
; CURRENT FILING DATE: 2002-04-16  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 260  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-123-904-260

Query Match 13.1%; Score 215; DB 9; Length 420;

Best Local Similarity 25.3%; Pred. No. 6.6e-13;

Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAK-IFVAGHRLVGSALVRKLOEQGFNTLVKLT-----HAELDLTR 45

Db 83 LSEKDRKRILITGGAGFVGSHTDKLMDGHEVTVVDNFTGKRKNVHEWHIGHENFELIN 142

QY 46 QADVESFFSQEPKPVYVILAAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103

Db 143 HDVVEPLYEVDQIYHLASPAS-----PPNYMNPITKLTNTIGTLMGLAKR 192

QY 104 ----LLFLGSSCIYPKFAPOPIPESSALLTASLEP--TNEWYAIAGIKTKQAYRIQHG 157

Db 193 VGARLLLASTSEVYGD--PEVHPQSEDYWGHNPIGPACRYDEGKRVAETMTCYAYMKQSG 250

QY 158 WDAISGMPTNLVGP-----NDNFHPENSHVLPALMRFPHEAKVNGAEVVVWGTGSPLERF 213

Db 251 VEVRVARIFTFGPRMMDNG-RVVSNFILQALQ-----GEPLTVYSGSGQTRAF 299

QY 214 LHVDDLADACVPLLDVRYSGLEHVNIGSGQEVTTIRELAEVLKVEVFGKLGW----- 265

Db 300 QYVSDLVNGLVALMNS-NVSSPVNLGNPEHTILEFAQLIKNLVSGSGSEIQFLSEAQDDP 358

QY 266 DCTKPDGTTPRKLMDSSKSLASLGWTPKVSRLDGLSQTQYDWYLNK 309

Db 359 QKRKPDIIKAKLM-----LGWEPVPLEEGLNKAHYFRKEL 395

## RESULT 11

US-10-140-470-260

; Sequence 260, Application US/10140470

; Publication No. US20030022331A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven



; APPLICANT: Smith,Victoria  
; APPLICANT: Stewart,Timothy A.  
; APPLICANT: Tumas,Daniel  
; APPLICANT: Watanabe,Colin K  
; APPLICANT: Wood,William  
; APPLICANT: Zhang,Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C160  
; CURRENT APPLICATION NUMBER: US/10/140,470  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 260  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-140-470-260

Query Match 13.1%; Score 215; DB 9; Length 420;  
Best Local Similarity 25.3%; Pred. No. 6.6e-13;  
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;  
QY 1 MSDKSAK-IFVAGHRLVGSIAVRKLEQEQFTNLVLT-----HAELDLTR 45  
DB 83 LSEKDRKRILITGGAGFVGSHTDLKMDGHEVTVVDNFFTKRKNVHWHGFENFELIN 142  
QY 46 QADVESFFSQEKPVYVILAAAKVGGIHANNYTPADFIGVNLQ-IQTNVIHSAVEHGVKK- 103  
DB 143 HDVVEPLYIEVDQIYHLASPAS-----PPNYMYPKIKTKNTIGTLNMLGLAKR 192  
QY 104 ----LFLGSSCIYKPAQPIPEPESALLTASLEP--TNEWYAIKTIAGIKTCQAYRIQHG 157  
DB 193 VGARLLLASTSEVIGD--PEVHPQSEDYMGVHNPIGPRACYDEGKRVAETMCTAYMKQEG 250  
QY 158 WDAISGMPTNLYGP-----NDNFHPENSHVLPALMRFRHEAKVNGABEVVWGTGSPREF 213  
DB 251 VEVVRVARIFTGFRMHMDNG-RVNSNFILOALQ-----GEPLTVYSGSQTRAF 299  
QY 214 LHVDLADACVFLLDYRSGLEHVNIGSGQEVTTIRELAELVKVEVPEGKLGW----- 265  
DB 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTTLEFAQLIKNLVSGSGEIOFLSEAQDDP 358  
QY 266 DCTKPDGTGPRKMDSSKSLASLGWTPKVSRLDGLSQTYDWYLVKNV 309  
DB 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGLNKAIHYPRKEL 395

RESULT 12  
US-10-175-746-260  
; Sequence 260, Application US/10175746  
; Publication No. US20030027270A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Godwin, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C353

; CURRENT APPLICATION NUMBER: US/10/175,746  
; CURRENT FILING DATE: 2002-06-19  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 260  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-175-746-260

Query Match 13.1%; Score 215; DB 9; Length 420;  
Best Local Similarity 25.3%; Pred. No. 6.6e-13;  
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;  
QY 1 MSDKSAK-IFVAGHRLVGSIAVRKLEQEQFTNLVLT-----HAELDLTR 45  
DB 83 LSEKDRKRILITGGAGFVGSHTDLKMDGHEVTVVDNFFTKRKNVHWHGFENFELIN 142  
QY 46 QADVESFFSQEKPVYVILAAAKVGGIHANNYTPADFIGVNLQ-IQTNVIHSAVEHGVKK- 103  
DB 143 HDVVEPLYIEVDQIYHLASPAS-----PPNYMYPKIKTKNTIGTLNMLGLAKR 192  
QY 104 ----LFLGSSCIYKPAQPIPEPESALLTASLEP--TNEWYAIKTIAGIKTCQAYRIQHG 157  
DB 193 VGARLLLASTSEVIGD--PEVHPQSEDYMGVHNPIGPRACYDEGKRVAETMCTAYMKQEG 250  
QY 158 WDAISGMPTNLYGP-----NDNFHPENSHVLPALMRFRHEAKVNGABEVVWGTGSPREF 213  
DB 251 VEVVRVARIFTGFRMHMDNG-RVNSNFILOALQ-----GEPLTVYSGSQTRAF 299  
QY 214 LHVDLADACVFLLDYRSGLEHVNIGSGQEVTTIRELAELVKVEVPEGKLGW----- 265  
DB 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTTLEFAQLIKNLVSGSGEIOFLSEAQDDP 358  
QY 266 DCTKPDGTGPRKMDSSKSLASLGWTPKVSRLDGLSQTYDWYLVKNV 309  
DB 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGLNKAIHYPRKEL 395

RESULT 13  
US-10-176-918-260  
; Sequence 260, Application US/10176918  
; Publication No. US20030027275A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Godwin, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C382  
; CURRENT APPLICATION NUMBER: US/10/176,918  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 260  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-176-918-260

Query Match 13.1%; Score 215; DB 9; Length 420;  
Best Local Similarity 25.3%; Pred. No. 6.6e-13;  
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAK-IFVAGHRLGVSIAVRKLEQGFTNLVLT-----HAELDLTR 45  
DB 83 LSEKDRKRILITGGAGFVGSHTDLKLMMDGHEVTVDNFFTGKRNVEHWIGHENFELIN 142  
QY 46 QADVESFSEKPVVYVILAAKVGGIHANNTYPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103  
DB 143 HDVVEPLYEVDQIYHLASPAS-----PPNYMYPKIKTKNTIGTGLNMLGLAKR 192  
QY 104 ----LFLGSSCTYPKPAQPIPESSALLTASLEP--TNEWYAIAGIKTKCOAYRIQHG 157  
DB 193 VGARLLLASTSEVYGD--PEVHQSEDYGHVNPIGPRACYDEGKRVAETMCYAYMKQEG 250  
QY 158 WDAISGMTNLYGP-----NDNFHPENSHVLPALMRRFHEAKVNGABEVVWGTGSPLEF 213  
DB 251 VEVVRVARIFNTFGPRMHMDG-RVVSNFLQALQ-----GEPLTVYSGSQTRAF 299  
QY 214 LHVDLDLADACVFLLDRLYSGLEHVNIGSGQEVITIRELAELVKEVVGEGKLGW----- 265  
DB 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTILEFAQLIKNLVSGSGEIQFLSEAQDDP 358  
QY 266 DCTKPDGTPRKLMDSKSLASLGWTPKVSRLDGLSQTVDWYLNKV 309  
DB 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGINKRAIHYFRKEL 395

## RESULT 14

US-10-176-921-260

; Sequence 260, Application US/10176921

; Publication No. US2003002726A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C288

; CURRENT APPLICATION NUMBER: US/10/176,921

; PRIOR FILING DATE: 2002-06-20

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 260

; LENGTH: 420

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-176-921-260

Query Match 13.1%; Score 215; DB 9; Length 420;  
Best Local Similarity 25.3%; Pred. No. 6.6e-13;  
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAK-IFVAGHRLGVSIAVRKLEQGFTNLVLT-----HAELDLTR 45  
DB 83 LSEKDRKRILITGGAGFVGSHTDLKLMMDGHEVTVDNFFTGKRNVEHWIGHENFELIN 142

QY 46 QADVESFSEKPVVYVILAAKVGGIHANNTYPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103  
DB 143 HDVVEPLYEVDQIYHLASPAS-----PPNYMYPKIKTKNTIGTGLNMLGLAKR 192  
QY 104 ----LFLGSSCTYPKPAQPIPESSALLTASLEP--TNEWYAIAGIKTKCOAYRIQHG 157  
DB 193 VGARLLLASTSEVYGD--PEVHQSEDYGHVNPIGPRACYDEGKRVAETMCYAYMKQEG 250  
QY 158 WDAISGMTNLYGP-----NDNFHPENSHVLPALMRRFHEAKVNGABEVVWGTGSPLEF 213  
DB 251 VEVVRVARIFNTFGPRMHMDG-RVVSNFLQALQ-----GEPLTVYSGSQTRAF 299  
QY 214 LHVDLDLADACVFLLDRLYSGLEHVNIGSGQEVITIRELAELVKEVVGEGKLGW----- 265  
DB 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTILEFAQLIKNLVSGSGEIQFLSEAQDDP 358  
QY 266 DCTKPDGTPRKLMDSKSLASLGWTPKVSRLDGLSQTVDWYLNKV 309  
DB 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGINKRAIHYFRKEL 395

## RESULT 15

US-10-227-884-240

; Sequence 240, Application US/10227884

; Publication No. US20030027988A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Desnoyers, Luc

; APPLICANT: Gerritsen, Mary

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, J. Christopher

; APPLICANT: Gurney, Austin L.

; APPLICANT: Smith, Victoria

; APPLICANT: Stephan, Jean-Philippe F.

; APPLICANT: Watanabe, Colin L.

; APPLICANT: Wood, William I.

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3530PIC79

; CURRENT APPLICATION NUMBER: US/10/227,884

; CURRENT FILING DATE: 2002-08-26

; PRIOR APPLICATION NUMBER: 10/119,480

; PRIOR FILING DATE: 2002-04-09

; PRIOR APPLICATION NUMBER: 60/059113

; PRIOR FILING DATE: 1997-09-17

; PRIOR APPLICATION NUMBER: 60/062287

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 60/063549

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/064103

; PRIOR FILING DATE: 1997-10-31

; PRIOR APPLICATION NUMBER: 60/069873

; PRIOR FILING DATE: 1997-12-17

; PRIOR APPLICATION NUMBER: 60/078910

; PRIOR FILING DATE: 1998-03-20

; PRIOR APPLICATION NUMBER: 60/079294

; PRIOR FILING DATE: 1998-03-25

; PRIOR APPLICATION NUMBER: 60/079656

; PRIOR FILING DATE: 1998-03-26

; PRIOR APPLICATION NUMBER: 60/079728

; PRIOR FILING DATE: 1998-03-27

; PRIOR APPLICATION NUMBER: 60/081819

; PRIOR FILING DATE: 1998-04-15

; PRIOR APPLICATION NUMBER: 60/081955

; PRIOR FILING DATE: 1998-04-15

; PRIOR APPLICATION NUMBER: 60/082804

; PRIOR FILING DATE: 1998-04-22

; PRIOR APPLICATION NUMBER: 60/084441

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: 60/085323

; PRIOR FILING DATE: 1998-05-13

; PRIOR APPLICATION NUMBER: 60/085579

1	PRIOR APPLICATION NUMBER: 60/106464	
2	PRIOR FILING DATE: 1998-10-30	
3	PRIOR APPLICATION NUMBER: 60/106905	
4	PRIOR FILING DATE: 1998-11-03	
5	PRIOR APPLICATION NUMBER: 60/108787	
6	PRIOR FILING DATE: 1998-11-17	
7	PRIOR APPLICATION NUMBER: 60/108801	
8	PRIOR FILING DATE: 1998-11-17	
9	PRIOR APPLICATION NUMBER: 60/108849	
10	PRIOR FILING DATE: 1998-11-18	
11	PRIOR APPLICATION NUMBER: 60/112422	
12	PRIOR FILING DATE: 1998-12-15	
13	PRIOR APPLICATION NUMBER: 60/113296	
14	PRIOR FILING DATE: 1998-12-22	
15	PRIOR APPLICATION NUMBER: 60/113605	
16	PRIOR FILING DATE: 1998-12-23	
17	PRIOR APPLICATION NUMBER: 60/113621	
18	PRIOR FILING DATE: 1998-12-23	
19	PRIOR APPLICATION NUMBER: 60/115558	
20	PRIOR FILING DATE: 1999-01-12	
21	PRIOR APPLICATION NUMBER: 60/115565	
22	PRIOR FILING DATE: 1999-01-12	
23	PRIOR APPLICATION NUMBER: 60/115733	
24	PRIOR FILING DATE: 1999-01-12	
25	PRIOR APPLICATION NUMBER: 60/119549	
26	PRIOR FILING DATE: 1999-02-10	
27	PRIOR APPLICATION NUMBER: 60/123618	
28	PRIOR FILING DATE: 1999-03-10	
29	PRIOR APPLICATION NUMBER: 60/125259	
30	PRIOR FILING DATE: 1999-03-19	
31	PRIOR APPLICATION NUMBER: 60/125775	
32	PRIOR FILING DATE: 1999-03-23	
33	PRIOR APPLICATION NUMBER: 60/126773	
34	PRIOR FILING DATE: 1999-03-29	
35	PRIOR APPLICATION NUMBER: 60/127987	
36	PRIOR FILING DATE: 1999-04-05	
37	PRIOR APPLICATION NUMBER: 60/130232	
38	PRIOR FILING DATE: 1999-04-21	
39	PRIOR APPLICATION NUMBER: 60/131022	
40	PRIOR FILING DATE: 1999-04-26	
41	PRIOR APPLICATION NUMBER: 60/131270	
42	PRIOR FILING DATE: 1999-04-27	
43	PRIOR APPLICATION NUMBER: 60/131291	
44	PRIOR FILING DATE: 1999-04-27	
45	PRIOR APPLICATION NUMBER: 60/131445	
46	PRIOR FILING DATE: 1999-04-28	
47	PRIOR APPLICATION NUMBER: 60/134287	
48	PRIOR FILING DATE: 1999-05-14	
49	PRIOR APPLICATION NUMBER: 60/140650	
50	PRIOR FILING DATE: 1999-06-22	
51	PRIOR APPLICATION NUMBER: 60/140723	
52	PRIOR FILING DATE: 1999-06-22	
53	PRIOR APPLICATION NUMBER: 60/141037	
54	PRIOR FILING DATE: 1999-06-23	
55	PRIOR APPLICATION NUMBER: 60/144758	
56	PRIOR FILING DATE: 1999-07-20	
57	PRIOR APPLICATION NUMBER: 60/145698	
58	PRIOR FILING DATE: 1999-07-26	
59	PRIOR APPLICATION NUMBER: 60/146222	
60	PRIOR FILING DATE: 1999-07-28	
61	PRIOR APPLICATION NUMBER: 60/146963	
62	PRIOR FILING DATE: 1999-08-03	
63	PRIOR APPLICATION NUMBER: 60/149320	
64	PRIOR FILING DATE: 1999-08-17	
65	PRIOR APPLICATION NUMBER: 60/149638	
66	PRIOR FILING DATE: 1999-08-17	
67	PRIOR APPLICATION NUMBER: 60/151733	
68	PRIOR FILING DATE: 1999-08-31	
69	PRIOR APPLICATION NUMBER: 60/164418	
70	PRIOR FILING DATE: 1999-11-09	
71	PRIOR APPLICATION NUMBER: 60/166361	
72	PRIOR FILING DATE: 1999-11-16	
73	PRIOR APPLICATION NUMBER: 60/169445	









Db 1070 GAGGTCTCAATCAAGAACTCGCCGAAATGGTGAAGAAAGTGGTTGGATTTCAGGGGCGAG 1129  
QY 787 CTTGGATGGGATTCAGTAAAGCAGATGGCACACCGAGGAAATCTTATGGACAGCTCAAAAG 846  
Db 1130 CTGACATGGGATCTTCTTAAGCCTGATGAACCTCCACGAAAGCTCATCGATAGCAGCAA 1189  
QY 847 CTCGGCTCTTTGGTTGGACACCTTAAGGTTTCTTTAGAGATGGTGTGAGCCAACTTAT 906  
Db 1190 CTTGCCAATCGGGGTGGGAAGCGAGAAATCCCTCAAGGAAGGATTGGCAGAGACTTAC 1249  
QY 907 GATTGGTATTGAAGAA 923  
Db 1250 AATGGTACTGTGAGAA 1266

## RESULT 3

US-09-894-844-49  
; Sequence 49, Application US/09894844  
; Patent No. US20020176873A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of  
; FILE REFERENCE: the M. tuberculosis Complex  
; CURRENT APPLICATION NUMBER: US/09/894,844  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: M. tuberculosis  
US-09-894-844-49

## Query Match

29.4%; Score 275.2; DB 9; Length 966;

Best Local Similarity 56.2%; Pred. No. 9.7e-81;

Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

QY 6 TGACAAATCTGCAAAATCTTCGTGCGGGTCAATCGTGGTTGGTGGATCTGCCATTGT 65  
Db 30 TGACCGGGCGGCGGGTCTACATCGCGGGCATCGCGCTGGTGGTTCGGCTGCT 89  
QY 66 CGCAAGCTTCAGGAACAAGTTTCAACCAATCTCGTTCTTAAACACACGCGGAGTTGA 125  
Db 90 ACGCAGTTTGGCGGCGGGGTTACCAACCTGCTGGTGGTCAACGCGCGGAGCTTGA 149  
QY 126 TCTCATCTGTCAGCGGATGTTGAATCCTCTCTTCTCAAGAGAACCGATTATGTAAT 185  
Db 150 TCTGACGGATCGGGCGCGGAGTTGCACTTCGTTCTCGAGTCGAGCGCGAGTCTCAT 209  
QY 186 CTTAGCAGCAGCTAAAGTTGGTGTATTCACGCTAAACACCTATCTCTGCTGATTTCAT 245  
Db 210 CGACGGGGCGGCGGGTGGCGGATCTTGGCCACAGCACCTACCGCGGATTTCT 269  
QY 246 TGGTGTCAATCTCCAGATTACAGCAATGTGATCCACTCTGCAATATGACACGGTGTGA 305  
Db 270 GTCCGAAACCTCCAGATCCAGTCAACCTGCTGGATCGCGCGTGGCGGGTGC 329  
QY 306 GAAGCTTCTCTCTTGGATCATCTCGCATTTACCTTAATTTGCTCCTCAGCAATTC 365  
Db 330 GGGGCTGCTGTTCTCGGCTCGTGGTCACTTACCGAAATCTCGCGCGGAGCTCC 389  
QY 366 TGAGTCTGCTTTCTTAACAGCATCGCTTGAACCAACTAATAGTGTGATGCTATTCTAA 425  
Db 390 GGAGAGCGGCTGCTCACCCTGCTGGTGGAGCGGCAACAGCGGCTACCGATCGCAA 449

QY 426 GATCGCTGGGATTAAGACTTTGTTCAGGCTTATAGGATTCAGCAGGATGGGATGCAATCTC 485  
Db 450 AATCGCGCGACTCTTGGCGTCCAGCGCGTGCAGCGCAACATATGCGCTCCGCTGGATCTC 509  
QY 486 TGGCATGCTACTAATCTCTATGGTCTTAATCACAATTTCCACCGGAGAAATCTCATGT 545  
Db 510 GCGGATGCCCAACCTGTACGGGCGAGCGCAACATTTTCGCGCTCCGGCTCGCATCT 569  
QY 546 GCTTCTCTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGGAGCGGAGAACTTGT 605  
Db 570 GCTGCGGCACTCATCCGCGCTATGACGAGGCGCAAGCCAGTGGCGCGCCCAACCTGAC 629  
QY 606 GGTGTGGGTACAGGTAGTCCGTTGAGGAGTCTTTCGATGTTGATGTTGGCTGATGC 665  
Db 630 CAACTGGGCGACCGGCGAGCGCGGAGTTGCTGCACTGCGACGACCTGGCGAGCGC 689  
QY 666 TTGTGTTTTCTTGTGATGATACAGCGGTTGGAGCATGTTAACTTGGAGTGGTCA 725  
Db 690 ATGCTGTATCTGTGGAACATTTGACGGGCGGACCCATGTCACGTTGGAAACCGGCAT 749  
QY 726 AGAAGTGAATATTAGAGAGTTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGAA 785  
Db 750 CGACACACCATCGCGAGATCGCGAGATGTTGCGCTCGCGGTAGGCTATAGCGCGCA 809  
QY 786 GCTTGGATGGATTCGACTAAGCCAGATGGCACACGAGGAAACTTATGACAGCTCAA 845  
Db 810 AACCCTGCGGATCCAAAGCAACCGGACGGAACACCAACCACTGCTGGATGTTTCGT 869  
QY 846 GCTCGCGCTTTGGGTTGGACACCTAAGGTTTCTCTTACAGATGGTCTGAGCCAACTTA 905  
Db 870 GCTACGGAGCGGATGCGGCTTCGATCGGCTGCGGAGCGCATCAGCGCGCGGT 929  
QY 905 TGATTGGTATTGAAGAATG 925  
Db 930 GCGGTGATTCGCGAGCAG 949

## RESULT 4

US-10-114-170-121/c  
; Sequence 121, Application US/10114170  
; Publication No. US20030023075A1  
; GENERAL INFORMATION:  
; APPLICANT: Blattner, Frederick R.  
; Blattner, Valerie  
; Perna, Nicole T.  
; Plunkett, Guy  
; Welch, Rod  
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170

FILING DATE: 01-Apr-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702

FILING DATE: 03-DEC-1999

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386



QY 562 AGAGGTTTCCAGGCGGAAAGTGAATGGAGCGGAGAAAGTTGTGTGGGTACAGGT 621  
DB 5550 CGCAAGACATGAGGCAAGATTAAGACCTTGGGTGCTTGTCTATATGGGAGCGGC 5491  
QY 622 AGTCCGTTGAGGAGTCTTGCATGTTGATTTGGCTGATGCTTGTCTTCTTGTG 681  
DB 5490 ACACCTACTCGAGACTTTTGTACAGTGAAGACTGCTCGAGCGCTGTCTTCTTCT 5431  
QY 682 GATCGATACAGCGGTTGGAGCATGTTAAACATTTGAAGTGGTCAAGAGTGAATTA 741  
DB 5430 AGCAATTTCCGAACGGAACACATTAACATAGGCTCGGGGGGAATAGTATCAT 5371  
QY 742 GAGTGGCTGAGTTGGTGAAGAGTGTGTGTTTGAAGGGAAGCTTGGATGGATTC 801  
DB 5370 GAACTAGCCACATCTGCTGCGGTGTGTGTTTGAAGGCGATATAGTCTTCGACACA 5311  
QY 802 ACTAAGCCAGATGGCACCGGAAACCTTATGGACAGCTCAAGCTCCGCTTTTGGT 861  
DB 5310 TCAAGCCGAGGAAACGCAACGAAAGCTTTTATAGCGAAGACTCGTGTGATGGT 5251  
QY 862 TGACACCTAAGGTTTCTCTTAGAGATGCTCTGAGCAAACTTATGATTGGTATTTGAAG 921  
DB 5250 TGGCGCCGAGACCTGCTCGAGCTGGAGTGGCAATCTATGAATCGTTTGTACG 5191  
QY 922 AATGT 926  
DB 5190 AATGT 5186

RESULT 6  
US-09-318-271-3  
; Sequence 3, Application US/09318271A  
; Patent No. US20020012979A1  
; GENERAL INFORMATION:  
; APPLICANT: Berry, Alan  
; APPLICANT: Running, Jeffrey A.  
; APPLICANT: Severson, David K.  
; APPLICANT: Burlingame, Richard P.  
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"  
; FILE REFERENCE: 3161-24  
; CURRENT APPLICATION NUMBER: US/09/318,271A  
; CURRENT FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 60/125,073  
; EARLIER FILING DATE: 1999-03-17  
; EARLIER APPLICATION NUMBER: 60/125,054  
; EARLIER FILING DATE: 1999-03-18  
; EARLIER APPLICATION NUMBER: 60/088,549  
; EARLIER FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(966)  
US-09-318-271-3

Query Match 28.0%; Score 262.2; DB 10; Length 966;  
Best Local Similarity 56.5%; Pred. No. 2.1e-76;  
Matches 526; Conservative 0; Mismatches 378; Indels 27; Gaps 1;

QY 21 AATCTTCGTCGGGTCATCGTGTGTTGGTTGGATCTGCATGTCCGCAAGCTTCAGGA 80  
DB 15 AGTTTTATGTGTGTCATCGCGGATGTTGGTTCGCGATCAGCGCGCAGCTCGAACA 74  
QY 81 ACAAGGTTTCCACCAATCTCGTCTTAAACACACACGCGGAGCTTGATCTCACTCGTCAAGC 140  
DB 75 GCGCGGTGATGTGGAATCTGTATTACGACCCGCGAGAGCTGAACCTCTGGACAGCGC 134  
QY 141 CGATGTTGAATCTCTTTTCTCAAGAGAAGCGAGTTTATGTAATCTTAGCAGCACTAA 200

DB 135 CCCCCTGTCATGATTTCTTTGCGAGGAAACGTATTACCCAGGCTCTATCTGCGCGCGCGCAA 194  
QY 201 AGTTCGTGATTTACCGCTAAACAACCTATCTCTGATTTTCATTTGGTGTCAATCTCCA 260  
DB 195 AGTGGCGGCATTTGTCACCAACAACCTATCCGGGGAATTTCACTACCAAGACATGAT 254  
QY 261 GATTCAGACCAATGTGATCCACTCTGTCATATAGCAGCGGTGGAAGAGCTTCTCTTCCT 320  
DB 255 GATTGAGAGCAACATCATTTACCGCGCGCATCAGAACGACGCTGAACAAATGCTGTTTCT 314  
QY 321 TGGATCATCTGTCATTTACCTTAATTTGCTCTCTAGCCAAATCTCTAGTCTGCTTTGTT 380  
DB 315 CGGATCGTCTGTCATCTACCCGAAATCTGGCAAAACAGCCGATGGCAGAAAGCGATGTT 374  
QY 381 AACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATTGCTAAGATCGCTGGGATTA 440  
DB 375 GCAGGGCAGCTGGAGCCGACATACGAGCCTTATGCTATTGCCAAATCGCCGGGATCAA 434  
QY 441 GACTTGTGAGGCTTATAGGATTCAGACGGAATGGGATGCAATCTCTGGCATGCTTAA 500  
DB 435 ACTGTGCGAATCATACACCGCAGTACGAGCGGATACCGCTCAGTCACTGCGACCAA 494  
QY 501 TCTCTATGCTCTAATGACAAATTTCCACCGGAGAAATCTCATGCTGCTTCTGCTCTTAT 560  
DB 495 CCTGTACGGGCCACAGCAACTTCCACCCGAGTAAATTCGCATGTGATCCAGCATTTGCT 554  
QY 561 GAGGAGGTTTCCAGGAGCGAAAGTGAATGGAGCGGAGGAAAGTTGTGCTGCGGTACAGG 620  
DB 555 GCGTCGCTTCCAGAGGCGACGCGACAGNATCGCCGAGCTGGTGGTATGGGCGAGCGG 614  
QY 621 TAGTCCGTTGAGGAGTCTTTCGATGTTGATGATTTGGCTGATGCTGTTGTTTCTTGTCT 680  
DB 615 TACACCGATGCGCGAATTTCTGCACTGATGATGCGCGCGGCGAGCATTCATGTCAT 674  
QY 681 GGATC-----GATACGCGGTTGGAGCATGTTAACT 713  
DB 675 GGAGCTGGCGCATGAAGTCTGGCTGGAGAACACCCAGCGCATGTTGTCGACATTAACGT 734  
QY 714 TGAAGTGTCAAGAGTGACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTCTTGG 773  
DB 735 CGGCGCGGCTTGACTGCATCTCCGCGAGCTGGCGCAACCATCGCCAAAGTGGTGG 794  
QY 774 TTTTGAAGGGAAGCTTGGATGGGATTCGACTAAGCCAGATGGCACACCGAGGAACTTAT 833  
DB 795 TTACAAAGCGCGGTGTTTTTGTATGCCAGCAACCGGATGCGACGCGCGCAAACTGCT 854  
QY 834 GGACAGCTCAAGCTCGGCTCTTTGGGTTGGACACCTAAGGTTTCTCTTAGAGATGCTCT 893  
DB 855 GGATGTGACGCGCTCGCATCAGCTTGGCTGGTATCAGGAAATCTCACTGGAAGCGGGCT 914  
QY 894 GAGCCAAACTTATGATTTGTTTGAAGAAT 924  
DB 915 TCCAGCACTTACCAGTGTTCCTTGAGAAT 945

RESULT 7  
US-09-734-569-15  
; Sequence 15, Application US/09734569  
; Patent No. US20020064816A1  
; GENERAL INFORMATION:  
; APPLICANT: Lerchl, Jens  
; APPLICANT: Renz, Andreas  
; APPLICANT: Ehrhardt, Thomas  
; APPLICANT: Reindl, Andreas  
; APPLICANT: Cirpus, Petra  
; APPLICANT: Bischoff, Friedrich  
; APPLICANT: Frank, Markus  
; APPLICANT: Freund, Annette  
; APPLICANT: Duwenig, Elke  
; APPLICANT: Schmidt, Ralf-Michael  
; APPLICANT: Reski, Ralf  
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved

```
; TITLE OF INVENTION: in the synthesis of carbohydrates
; FILE REFERENCE: BASF-NAB-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 15
; LENGTH: 701
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(487)
; OTHER INFORMATION: 20_ppprot1_064_d07
US-09-734-569-15

Query Match      22.8%; Score 213.6; DB 10; Length 701;
Best Local Similarity 65.5%; Pred. No. 2.5e-60;
Matches 312; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 448 CAGGCTTATAGATTTCAGCAGGATGGATGCAATCTCTGGCATGCTTACTTACTTATCTTAT 507
Db      |||||
QY 508 GGTCCCTAATGACAAATTCACCGGAGAAATCTCATGTGCTTCCCTCTTATGAGGAG 567
Db      |||||
QY 62  GGTCCCCACGACAAATTCATCCGAGAACTCCACGCTTTCGCCAGCTTGTATGACAGCG 121
Db      |||||
QY 568 TTCCAGGCGGCGAAATGGAATGAGCGGAGGAAGTTGTGGTGTGGGGTACAGGTATGCG 627
Db      |||||
QY 122 TTTCAGGAGGCTAAGGTGAACGGCGCTAAGGAAGTGGTGTGGGGATCAGGTTCCCA 181
Db      |||||
QY 628 TTGAGGGAGTCTTTCATGTTGATGATTTGGTGTGATGCTTGTGTTTCTTGTGATGCA 687
Db      |||||
QY 182 TTCCGTGAGTTCTTCACGTGACGACTTGGCAGAGGCAACAGTATTTCTGCTGCAGAT 241
Db      |||||
QY 688 TACAGCGGTGTGAGCATGTTAATGATTTGAAGTGTGCAAGAGTGAATTTAGAGAGTTG 747
Db      |||||
QY 242 TACTCCGCGATGAGCATGTCAATGCGGAGTGGCTCTGAGGCTCTCAATCAAGGAATC 301
Db      |||||
QY 748 GCTGAGTTGGTGAAGAGGTTGTGTTTGAAGGGAAGCTTTGGATGGGATTCGCACTAAG 807
Db      |||||
QY 302 GCGGAAATGGTGAAGAGTGGTGGATTTTCAAGGGGCACTGCATGGGATCTTCTAAG 361
Db      |||||
QY 808 CCAGATGGCACCCGAGGAACTTATGGAAGCTCAAGCTCGGCTTGTGGGTTGGACA 867
Db      |||||
QY 362 CCTGATGGAATCCACGAAAGCTCATCGATAGCAGCAAACTTGCACACATGGGGTGGCA 421
Db      |||||
QY 868 CCTAAGGTTTCTCTTAGAGATGCTCTGAGCCAACTTATGATGTTGTTTGAAGAA 923
Db      |||||
QY 422 GCGAGAAATCCCTCAAGGAAGATTGGCAGAGACTTACAAATGGTACTGTGAGAA 477
Db      |||||

RESULT 8
US-09-962-805-2
; Sequence 2, Application US/0962805
; Patent No. US20020058313A1
; GENERAL INFORMATION:
; APPLICANT: RENKONEN, Risto
; APPLICANT: MATILA, Pirkko
; APPLICANT: HIRVAS, Laura
; APPLICANT: HORTLING, Solveig
; APPLICANT: KALLIOINEN, Tuula
; APPLICANT: KAURANEN, Sirka-Liisa
; APPLICANT: JAEVINEN, Nina
; APPLICANT: MAEKI, Minna
; APPLICANT: NIITYMAEKI, Jaana
; APPLICANT: RAEBINAE, Jarkko
; TITLE OF INVENTION: USE OF RECOMBINANT ENZYMES FOR PREPARING GDP-L-FUCOSE AND FUCOSYL
; FILE REFERENCE: 2242/50463
```

```
; CURRENT APPLICATION NUMBER: US/09/962,805
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FI 20002114
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-09-962-805-2

Query Match      15.6%; Score 145.8; DB 10; Length 933;
Best Local Similarity 50.4%; Pred. No. 1.1e-37;
Matches 411; Conservative 0; Mismatches 397; Indels 7; Gaps 2;

QY 117 CGAGCTTCACTCTCAAGCGGATGTTGAATCTTCTTTCTCAAGAGAACCGAGT 176
Db      |||||
QY 111 CGAATTGTTATTTGTTGGATAAAGACAACGTTTCAAGCCTATTTTGAAGAATAACAGCCTAC 170
Db      |||||
QY 177 TTATGTAATCCTAGCAGCAGCTAAAGTTGGTGTATTTCACGCTAACACACCTATCTCTGC 236
Db      |||||
QY 171 AGGCATATCATTTGTCGGGAGAGTGGGGGCAATTGTGGAACATGAAGATCTTTTC 230
Db      |||||
QY 237 TGATTTCATTTGGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCA 296
Db      |||||
QY 231 AACITACATGTTGAGAAATTTACTCATGGGTTTGTATCTTTTCTAGCGCTTTAGATTT 290
Db      |||||
QY 297 CGGTGTGAAGAGCTTCTTCTTCTTGGATCATCTCTGATATTAACCTTAATTTGCTCTCA 356
Db      |||||
QY 291 GGGCGTGAAGAAAGCCATTAATCTAGCGAGCTTCTTGCGCTTATCTTAATAGCGCCCTAA 350
Db      |||||
QY 357 GCCAATTCCTCAGTCTGCTTTGTTAAACAGCATCGTTGAACCACTAATGATGATGCTATGC 416
Db      |||||
QY 351 CCCTTAAAGAGAGCGAATTTATTTGAACCGCTCTTTAGAACCAAGATGAAGCTACGC 410
Db      |||||
QY 417 TATTGCTAAGATCGCTGGGATTAAGACTTGTACGGCTTTATAGGATTCACGACGGATGGGA 476
Db      |||||
QY 411 TTTAGCCAACTCTCTGTGATGAAGTATTCGAAATACGTGAGCGCTGAAAAGGGGTTT 470
Db      |||||
QY 477 TGCATCTCTGGCATGCTTACTTAATCTCTATGCTCTTAATGACAAATTTCCACCGGAGAA 536
Db      |||||
QY 471 TTATAAACTCTAGTGGCTTGTAACTTTATGCGGAGTTTGACAAAGTTTGAAGAAAGAT 530
Db      |||||
QY 537 TTCTCATGTGCTCTCTGCTTTATGAGGAGGTTCCACGAGCGAAAGTGAATGGAGCGGA 596
Db      |||||
QY 531 AGCGCATATGATCAAGGGCTTTATTTGATAGATGACACCCCTAAATTTAAATAATGAAAA 590
Db      |||||
QY 597 GGAAGTTGTGTGGGTACAGGTAGTCTCGTTGAGGGAGTTCTTTGCAATGTTGATGATTT 656
Db      |||||
QY 591 AAATTTTGGCATGTGGGCGATGGCAGCGCCAGAGAGAGTATCTAAACGCTAAAGATTT 650
Db      |||||
QY 657 GGC-----TGATGCTGTGTTTCTTCTGCTGATGATACAGCGGTTGGAGCATGTTTAAAC 711
Db      |||||
QY 651 AGCCAGATTCTCGCTCTCGCTTATGAGAAATATCGCTCAAAATGCTCGCTGATG--AAT 708
Db      |||||
QY 712 ATTGGAAGTGTCAAGAAGTGAATTTAGAGAGTTGGCTGAGTTGGTGAAGAGAGTTGTT 771
Db      |||||
QY 709 GTCGCTCTGGAGTGGATTACAGCATTTGAAGAGTATTACGAAAGATTCGCTCAGGTTTAA 768
Db      |||||
QY 772 GGTTTTGAAGGGAAGCTTTGATGGGATTTGCACTAAGCCAGATGGCACACCGAGGAACTT 831
Db      |||||
QY 769 GACTATAAGGGCGTGTGTTGTAAGAGATTTCATCCAAACCGAGTGGGCGATGCAACAAAGCTT 828
Db      |||||
QY 832 ATGACAGCTCAAGAGCTCGGCTCTTTGGGTTGGACACCTAAGGTTTCTCTTAGAGATGCT 891
Db      |||||
QY 829 ATGGATATTTCCAAACAAAGGCTTTAAATGGGAATTTAGAAATCCCTTTTAGAGCAGGCG 888
Db      |||||
QY 892 CTGAGCCAACTTATGATGTTGTTTGAAGAAATGT 926
Db      |||||
QY 889 ATCAAGAGCTTATGAGTATTATTGAGCTTTT 923
Db      |||||
```

## RESULT 9

US-09-318-271-5  
; Sequence 5, Application US/09318271A  
; Patent No. US20020012979A1  
; GENERAL INFORMATION:  
; APPLICANT: Berry, Alan  
; APPLICANT: Running, Jeffrey A.  
; APPLICANT: Severson, David K.  
; APPLICANT: Burlingame, Richard P.  
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"  
; FILE REFERENCE: 3161-24  
; CURRENT APPLICATION NUMBER: US/09/318,271A  
; CURRENT FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 60/125,073  
; EARLIER FILING DATE: 1999-03-17  
; EARLIER APPLICATION NUMBER: 60/125,054  
; EARLIER FILING DATE: 1999-03-18  
; EARLIER APPLICATION NUMBER: 60/088,549  
; EARLIER FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 5:  
; TYPE: DNA  
; LENGTH: 1340  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (75)..(1040)  
US-09-318-271-5

Query Match 10.9%; Score 101.6; DB 10; Length 1340;

Best Local Similarity 48.7%; Pred. No. 7.4e-23;

Matches 340; Conservative 0; Mismatches 349; Indels 9; Gaps 2;

QY 124 GATCTACCTGTCAGCCGATGTGAATCCTCTCTTCTTCTCAAGAGAGCCAGTTTATGTA 183  
DB 213 GATCTACCGGATACAGACAGCCCGCCCTGTGTGAGAGGTCCAAACCCACACAGTC 272  
QY 184 ATCTAGACAGCAGTAAAGTTGGTGTATTCAGCTTAAACACACCTATCTCTGATTTTC 243  
DB 273 ATCCATCTTGTCGAATGTGGGGGCTGTTCGGAATATCAATACATTTTGGACTTC 332  
QY 244 ATTGGTGCATCTCAGATTCAGACCAATGTATCCATCTGCATATAGACAGCGGTG 303  
DB 333 TGGAGGAAAAGTGCACATGAACCAACGCTCTGCCTCGCCCTTTGAGGTGGGGCC 392  
QY 304 AAGAGCTTCTCTCTTGGATCATCTGCAATTTACCTTAATTTGCTCTCAGCAAT 363  
DB 393 CGCAAGGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452  
QY 364 CTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423  
DB 453 GATGAGACCATGATCCACATGGGCTCCCAACAGCAATTTGGGTACTGATGCC 512  
QY 424 AAGATCGTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGGATGGGATGCAATC 483  
DB 513 AAGAGGATGATCGAGTGTGAGAAACAGGGCTACTTCCAGCAGTACGGCTGCACCTTCACC 572  
QY 484 TCTGGATCGCTACTTAATCTCTATGCTCTATATGCAATTTCCACCGGAGATTTCTCAT 543  
DB 573 GCTGTCATCCCCACCAACGTTTTCGGGGCCCCAGCAACTTCAACATCAGGATGCCAC 632  
QY 544 GTGCTTCTGCTGCTTATGAGGAGTTTCCACGAGGGGAAAGTGAATGAGCGGAGGAGTT 603  
DB 633 GTGCTGCTGCTGCTCATCACAGGTGACCTGGCCAGAGAGCGGCTCGGCC---TG 689  
QY 604 GTGCTGTTGGGTACAGGTAGTCCGTTGAGGAGTCTTTCATATGATGATTTGGCTGAT 663  
DB 690 ACGGTGTGGGTACAGGAAATCCGCGGAGCGGATTCATATCTCTGCTGAGCTGGCCAG 749  
QY 664 GCTTGTGTTTCTTCTGCTGATCGATACAGCGGGTGGAGCATGTTA-----ACATGGA 717  
DB 750 CTCCTTATCTGGGTCTCGGGAGTACAAATGAAGTGGAGGCCCATCATCTCTCCGTGGGC 809

QY 718 AGTGTCAAGAGTCACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGTTT 777  
DB 810 GAGGAAGATGAGTCTCCATCAAGAGGAGCGGCGGTGCTGAGGCCATGAGCTTC 869  
QY 778 GAAGGGAAGCTTGGATGGATTCCTAAGCCAGATGG 815  
DB 870 CATGGGAAGTCACCTTTGATACAAACCAAGTCGGATGG 907

## RESULT 10

US-09-294-093B-2058/C  
; Sequence 2058, Application US/09294093B  
; Patent No. US20010051335A1  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath, V.  
; APPLICANT: Ito, Laura, Y.  
; APPLICANT: Sherman, Bradley, K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
; FILE REFERENCE: PL-0009 US  
; CURRENT APPLICATION NUMBER: US/09/294,093B  
; CURRENT FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: 60/082,567  
; PRIOR FILING DATE: April 21, 1998  
; NUMBER OF SEQ ID NOS: 6207  
; SOFTWARE: PERL Program  
; SEQ ID NO 2058  
; LENGTH: 282  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700345468H1  
; NAME/KEY: unsure  
; LOCATION: 2, 6, 50  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-294-093B-2058

Query Match 6.4%; Score 59.8; DB 10; Length 282;

Best Local Similarity 56.8%; Pred. No. 2.2e-09;

Matches 151; Conservative 0; Mismatches 108; Indels 7; Gaps 2;

QY 23 TCTTCGTGCGGGTCAATCGTGGTTGGTGGATCTGCCATTCGCGCAAGCTTCAGGAAC 82  
DB 279 TCTTCGTGCTCGGGCGCAACGATCTGTGCGTTAAACCATCTGTGTGCTCCTCATGC 220  
QY 83 AGGTTTCAACCAATCTGTTCTTAAACACAGCCGAGCTTCATCTCACTCGTCAAGCG 142  
DB 219 TTGTTTCACTGCGTGTGCGCCACACCCAGCTTGACCTCACCCGCGGCTA 160  
QY 143 ATGTTGAATCCTTCTTTTCTCAA-----GAGAAGCCAGTTTATGTAACTCTAGCAGAG 196  
DB 159 ACTTTGAGGCTTCTTTCACCGCGGAGCGCGCGGCTATGTGTTCTCATGGCT 100  
QY 197 CTAAAGTTGGTGGTATCAOGCTAAACACCTATCTCTGCTGATTTCATT-GGTGTCAAT 255  
DB 99 CGAGTTTGAAGATCCATGCCAACTCGGCTTCCCCACCGACTTCATNCACAGTGAAC 40  
QY 256 CTCCAGATTCAGACCAATGTATCCA 281  
DB 39 CTCTAGATCTAGATCAACTTTGTGCA 14

## RESULT 11

US-10-025-380-296/C  
; Sequence 296, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy





QY 124 GATCTCACTCGTCAAGCCGATGTTGAATCGTCTCTTTCTCAAGAGAGCCAGTTTATGTA 183  
Db |||||  
QY 416 GATCTCAAGGATACAGCAGACAGCCCGCCCTGTTTGAAGAGGTCCAACCTCACACAGTC 357  
Db |||||  
QY 184 ATCTAGCAGCAGCTAAAGTTGGTGTATTTCAGCTTAACAAACACCTTATCTCTGATTTC 243  
Db |||||  
QY 356 ATCCATCTTGTGCAATGTTGGGGCCCTGTTCCGGATATCAATACAAATTGGACTTC 297  
Db |||||  
QY 244 ATTGGTCAATCTCAGATTGAGCAATGATCCACTCTGCAATATGAGCAGCGGTG 303  
Db |||||  
QY 296 TGGAGGAAAAACGTGCACATGACGACAAAGCTCTGCATCGGCCCTTCGAGGTGGCGCC 237  
Db |||||  
QY 304 AAGAAGCTTCTCTCTTGTGATCATCTGCAATTAACCTAAATTTGCTCTCAGCCAAAT 363  
Db |||||  
QY 236 CGCAAGGTGGTGTCTCTGCTGCTCACTGTATCTTCCCTGACAAGACGACCTACCCGATA 177  
Db |||||  
QY 364 CTGAGTCTGCTTTGTTAAACAGCATCGTCTGAACCAACTAAAGAGTGTATGTTGCT 423  
Db |||||  
QY 176 GATGAGACCATGATCCATATGAGCTCTCCCAACAGCAATTTGGGTACTGATGCC 117  
Db |||||  
QY 424 AAGATCGCTGGGATTGAAGCTTTGTCAAGGCTTATAGGATTCAGCAGGATGGGATGCAATC 483  
Db |||||  
QY 116 AAGAGGATGATGACGTGCAGAACAGGGCTACTTCCAGCAGTACGGCTGCACCTTCACC 57  
Db |||||  
QY 484 TCTGGATCGCTACTAATCTCTATGTCTCTATGCAATTTCCACCCGGAGAAAT 537  
Db |||||  
QY 56 GCTGTATCCCCACCAACGTCTTCGGGGCCCCACGACAACTTCAACATCGAGGAT 3

## RESULT 14

US-09-815-343-253

; Sequence 253, Application US/09815343

; Patent No. US20010055596A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun E.

; APPLICANT: King, Gordon E.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE REFERENCE: 210121.504

; CURRENT APPLICATION NUMBER: US/09/815,343

; CURRENT FILING DATE: 2001-03-22

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 253

; LENGTH: 656

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(656)

; OTHER INFORMATION: n = A,T,C or G

US-09-815-343-253

Query Match 4.8%; Score 44.8; DB 10; Length 656;  
Best Local Similarity 46.7%; Pred. No. 0.00037;  
Matches 142; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 124 GATCTCACTCGTCAAGCCGATGTTGAATCGTCTCTTTCTCAAGAGAGCCAGTTTATGTA 183  
Db |||||  
QY 144 GATCTCAGGATACAGCAGACAGCCCGCCCTGTTTGAAGAGGTCCAACCCACACAGTC 203  
Db |||||  
QY 184 ATCTAGCAGCAGCTAAAGTTGGTGTATTTCAGCTTAACAAACACCTTATCTGCTGATTTC 243  
Db |||||  
QY 204 ATCCATCTTGTGCAATGTTGGGGCCCTGTTCCGGATATCAATACAAATTGGACTTC 263  
Db |||||  
QY 244 ATTGGTCAATCTCAGATTTCAGCAATGATCCACTCTGCAATATGAGCAGCGGTG 303  
Db |||||  
QY 264 TGGAGGAAAAACGTGCACATGAACGACAGCTCTGCACTCGGCCCTTGAAGTGGCGCC 323  
Db |||||  
QY 304 AAGAAGCTTCTCTCTTGTGATCATCTGCAATTAACCTAAATTTGCTCTCAGCCAAAT 363  
Db |||||  
QY 324 CGCAAGGTGGTGTCTGCTGCTCCACTGTATCTTCCCTGACAAGACGACCTTACCCGATA 383

QY 364 CTTGAGTCTGCTTTGTTAAACAGCATCGTCTGTAACCAACTAATGAGTGGTATGCTATTGCT 423  
Db |||||  
QY 384 GATGAGACCATGATCCCAATGGGCTCCCCACACAGCAATTTGGGTACTCTGATGCC 443  
Db |||||  
QY 424 AAGA 427  
Db |||||  
QY 444 AAGA 447

## RESULT 15

US-09-770-445-638/C

; Sequence 638, Application US/09770445

; Patent No. US20020023281A1

; GENERAL INFORMATION:

; APPLICANT: Goriach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Krickler, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; FILE REFERENCE: 2023US (PARA-012PRV)

; CURRENT APPLICATION NUMBER: US/09/770,445

; PRIOR FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: US 60/178,472

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 638

; LENGTH: 846

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-445-638

Query Match 4.7%; Score 43.8; DB 10; Length 846;  
Best Local Similarity 51.2%; Pred. No. 0.00095;  
Matches 155; Conservative 0; Mismatches 142; Indels 6; Gaps 2;

QY 595 GAGGAAGTTGCTGGGTGACAGTAGTCCGTTGAGGAGTCTTGGCATGTTGATGAT 654  
Db |||||  
QY 603 GAGCCATTGACTGTTTACGTTGATGGGAAGCAGACAGAGGATTTCCAAATTTGTTCTGAT 544  
Db |||||  
QY 655 TTGGCTGATGCTGTTGTTTTCTTGTGATCGATACAGCGGTTGGAGCATGTTAAACATT 714  
Db |||||  
QY 543 CTGGTTGAAGGTTTGAAGACTGATGGAAGGAGAACATG---TCGGCCCATTCACCTC 487  
Db |||||  
QY 715 GGAAGTGGTCAAGAAAGTGAATATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGT 774  
Db |||||  
QY 486 GGTAAACCTTGGTGAATTCAGATGCTCGAGCTCGCTAAGTGGTCCAGAGACAAATTGAT 427  
Db |||||  
QY 775 TTTGAAGGGAAGCTTGGATGGGATTCACCTAAGCCAGATGGCACACCGAGGAACTTATG 834  
Db |||||  
QY 426 CGGAATGCAAAACATAGAGTTTCAGACCCAAACACAGAGACGACCTCACAGAGAAAGCCT 367  
Db |||||  
QY 835 GACAGCTCAAGCTC---CGGTCTTTGGGTTGACACCTAAGGTTTCTCTTAGAGATGTT 891  
Db |||||  
QY 366 GACATCAAAAGCCCAAGAGCTTTTAGGTTGGGAACCAAGAGTCTCTCTTCTGTCAGGGA 307

Oy 892 CTG 894  
|||  
Db 306 CTG 304

Search completed: June 3, 2003, 05:01:27  
Job time : 194 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 04:45:36 ; Search time 65 Seconds  
(without alignments)  
989.026 Million cell updates/sec

Title: US-10-089-014-1

Perfect score: 1646

Sequence: 1 MSDKSAKIFVAGHGLVGS.....SLRDGLSQTYDWYLNKNCNR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_muc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_ivirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1646	100.0	312	10	O49213 arabidopsis
2	1478	89.8	328	10	Q91MU0 arabidopsis
3	971	59.0	339	2	O86293 mycobacteri
4	971	59.0	339	2	O86298 mycobacteri
5	969	58.9	321	2	Q9RCB2 yersinia ps
6	956.5	58.1	321	2	Q56873 yersinia en
7	955	58.0	322	16	P71791 mycobacteri
8	944.5	57.4	321	16	Q8XGS8 salmonella
9	944.5	57.4	321	16	Q8X4R4 escherichia
10	928.5	56.4	321	2	Q9F7A3 salmonella
11	915.5	55.6	321	16	Q9S5F8 escherichia
12	915.5	55.6	323	2	O85340 escherichia
13	895.5	54.4	331	16	Q8U6L7 agrobacteri
14	893	54.3	360	2	Q8VU14 bacteroides
15	882	53.6	317	12	Q84611 paramecium
16	877.5	53.3	309	16	Q97H34 clostridium

17	848	51.5	322	16	Q98AU4 rhizobium 1
18	773.5	47.0	332	2	Q9JN55 coxiella bu
19	686	41.7	346	16	Q9PMM9 campylobact
20	672	40.8	289	2	Q9AQ09 bradyrhizob
21	621	37.7	208	2	Q9Z612 escherichia
22	596.5	36.2	246	16	Q9XDD7 brucella me
23	593.5	36.1	307	2	Q46720 escherichia
24	584.5	35.5	307	2	Q93Q29 salmonella
25	567.5	34.5	314	16	Q8YMW0 anabaena sp
26	565.5	34.4	312	16	P72585 synechocyst
27	547.5	33.3	308	2	O87152 vibrio chol
28	547.5	33.3	308	2	O51840 vibrio chol
29	540.5	32.8	307	2	Q8VQ41 escherichia
30	538.5	32.7	307	2	Q9F119 escherichia
31	531.5	32.3	308	2	O34227 vibrio chol
32	531.5	32.3	221	2	Q9R6T5 synechococ
33	531	32.3	287	2	Q9ZAY6 anabaena sp
34	505	30.7	310	16	O24886 helicobacte
35	502	30.5	310	2	O84974 helicobacte
36	495	30.1	310	16	Q9ZN21 helicobacte
37	437.5	26.6	321	5	Q9W1X8 drosophila
38	408	24.8	306	10	Q9SMD7 laminaria d
39	365.5	22.2	315	5	Q21632 caenorhabdi
40	345.5	21.0	114	16	Q8YBP7 brucella me
41	313.5	19.0	325	2	Q9FB21 streptomyce
42	273.5	16.6	309	17	Q8TXF0 methanopyru
43	270.5	16.4	310	17	Q9UXJ4 sulfolobus
44	266	16.2	309	16	Q9WYX9 thermotoga
45	264.5	16.1	310	16	Q55412 synechocyst

#### ALIGNMENTS

RESULT 1

O49213 ID O49213 PRELIMINARY; PRT; 312 AA.

AC O49213; DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase  
DE (Fragment).  
GN GER1 OR T18K17.8 OR ATFX.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA ECOTYPE;  
RA Bonin C.P., Potter I., Vanzin G.F., Reiter W.-D.;  
RT "A bifunctional epimerase-reductase completes the de novo synthesis of  
GDP-4-fucose in Arabidopsis.";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RT "Arabidopsis thaliana chromosome I BAC T18K17 genomic sequence.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RA Nakayama K., Maeda Y., Wang X., Jigami Y.;  
RT "Expression of the genes for GDP-fucose synthesis in yeast  
Saccharomyces cerevisiae.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF045286; AAC02703.2;  
DR EMBL; AC010556; AAG52124.1;  
DR EMBL; AB034806; BAA95670.1;

DR HSSP; P32055; 1BSV.  
 FT NON TER 312 312  
 SQ SEQUENCE 312 AA; 34477 MW; 4F5E77D9FA492C0B CRC64;

Query Match 100.0%; Score 1646; DB 10; Length 312;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-138;  
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLAETHAELDLTROADVSEFSSQEKPVY 60  
 DB 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLAETHAELDLTROADVSEFSSQEKPVY 60

QY 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYPKFAPQ 120  
 DB 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYPKFAPQ 120

QY 121 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 180  
 DB 121 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 180

QY 181 HVLPALMRPFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 240  
 DB 181 HVLPALMRPFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 240

QY 241 GQEVITIRELAELVKEVVGEGKLGWDCITKPDGTPRKLMDSSKSLASLGWTPKVSRLRDLG 300  
 DB 241 GQEVITIRELAELVKEVVGEGKLGWDCITKPDGTPRKLMDSSKSLASLGWTPKVSRLRDLG 300

QY 301 TYDWYLNKVCNR 312  
 DB 301 TYDWYLNKVCNR 312

RESULT 2  
 Q9LMU0 PRELIMINARY; PRT; 328 AA.

ID 09LMU0  
 AC 09LMU0  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE F2H15.12 protein (putative GDP-L-fucose synthetase).  
 GN F2H15.12 OR ATG17890.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLOMBIA;  
 RA Sakano H., Liu S.X., Btgu P., Lee J.M., Lenz C., Pham P., Toriumi M.,  
 RA Yu G., Chan A., Chung M., Goldsmith A., Liu A., Smith A., Vaysberg M.,  
 RA Altati H., Brooks S., Buehler E., Chao Q., Conn L., Conway A.B.,  
 RA Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M.,  
 RA Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,  
 RA Federspiel N.A., Theologis A.;  
 RT "The sequence of BAC F2H15 from Arabidopsis thaliana chromosome 1.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayaishiraki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full length cDNA of gene Atg17890 (GI:15220878).";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC034106; AAF97269.1; -;  
 DR EMBL; AY063880; AAL36236.1; -;  
 DR HSSP; P32055; 1BSV.

DR InterPro; IPR002198; ADH short.  
 DR PROSITE; PS00061; ADH SHORT; UNKNOWN 1.  
 SQ SEQUENCE 328 AA; 36227 MW; 31C4479E9C5BIAC6 CRC64;

Query Match 89.8%; Score 1478; DB 10; Length 328;  
 Best Local Similarity 88.5%; Pred. No. 4.5e-123;  
 Matches 276; Conservative 19; Mismatches 17; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLAETHAELDLTROADVSEFSSQEKPVY 60  
 DB 15 MLEKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLAETHAELDLTROADVSEFSSQEKPVY 74

QY 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYPKFAPQ 120  
 DB 75 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYPKFAPQ 134

QY 121 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 180  
 DB 135 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 194

QY 181 HVLPALMRPFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 240  
 DB 195 HVLPALMRPFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 254

QY 241 GQEVITIRELAELVKEVVGEGKLGWDCITKPDGTPRKLMDSSKSLASLGWTPKVSRLRDLG 300  
 DB 255 GQEVITIRELAELVKEVVGEGKLGWDCITKPDGTPRKLMDSSKSLASLGWTPKVSRLRDLG 314

QY 301 TYDWYLNKVCNR 312  
 DB 315 TYEWYLNKVCNR 326

RESULT 3  
 O86293 PRELIMINARY; PRT; 339 AA.

ID O86293  
 AC O86293  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE GsbB protein.  
 GN GSB.  
 OS Mycobacterium paratuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1770;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bull T.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,  
 RA Hermon-Taylor J.;  
 RT "A low G C content element in Mycobacterium avium subsp.  
 RT paratuberculosis and M. avium subsp. silvaticum with homologous genes  
 RT in M. tuberculosis";  
 RL Microbiology 144:3413-3423 (1997).  
 RL EMBL; AJ223833; CAAL1576.1; -;  
 DR HSSP; P32055; 1BSV.  
 SQ SEQUENCE 339 AA; 37188 MW; A18AEBB191435B9C CRC64;

Query Match 59.0%; Score 971; DB 2; Length 339;  
 Best Local Similarity 59.2%; Pred. No. 5.5e-78;  
 Matches 181; Conservative 50; Mismatches 75; Indels 0; Gaps 0;

QY 3 DKSARKIFVAGHRLVGSIVRKLQEQGFTNLVLAETHAELDLTROADVSEFSSQEKPVY 62  
 DB 28 DRATPVYTAGHRLVGSIVRKLQEQGFTNLVLAETHAELDLTROADVSEFSSQEKPVY 87

QY 63 LAAKVGGIHANNYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYPKFAPQ 122  
 DB 88 DAAARVGGIMANNYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYPKFAPQ 147

```

QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDAISGMPTNLYGPDNDNPHFPHSV 182
DB 148 ESALLTGPLEPTNDAYAIKAGILQVQAVRRQYGLAWISAMPTNLYGPDNFSGSHL 207
QY 183 LPALMRRPHEAKVNGAEVWVGTSPLREFLHVDDLDADACVFLDDRYSGLEHVNIGSQ 242
DB 208 LPALIRRYEEAKAGGAEEVNTGCTPRRELHVDDLDASACFLLEHFDGPNHVNVTGV 267
QY 243 EYTIRELAEVLKVEVGFEGKLGWCTKPDGTPRKLMDSKSLASLGWTPKVSURDGLSOTY 302
DB 268 DHSISEIADMVATAVGIETRWDPKPDGTPRKLLDVLSALRELGNRPRIALKDGDATV 327
QY 303 DWYLNK 308
DB 328 SWYRTN 333

RESULT 4
ID O86298 PRELIMINARY; PRT; 339 AA.
AC O86298;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GSBB protein.
GN GSBB OR MERA.
OS Mycobacterium avium subsp. silvaticum, and
OS Mycobacterium avium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=44282, 1764;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.avium subsp. silvaticum;
RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,
RA Hermon-Taylor J.;
RA "A low G+C content element in Mycobacterium avium subsp.
RT paratuberculosis and M. avium subsp. silvaticum with homologous genes
RT in M. tuberculosis.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RV [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.avium; STRAIN=2151;
RA Eckstein T.M., Lambert M.L., Brennan P.J., Belisle J.T., Inamine J.M.;
RA "Identification of a gene cluster involved in glycopeptidolipid
RT biosynthesis and of a gene cluster encoding daunorubicin resistance in
RT two strains of Mycobacterium avium serovar 2.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ223832; CAAL1572.1; -
DR EMBL; AF143772; RAD44220.1; -
DR EMBL; AF125999; AAD20374.1; -
DR HSSP; P32055; IBSV.
SQ SEQUENCE 339 AA; 37160 MW; 701B483E2CD3E7AF CRC64;

Query Match 59.0%; Score 971; DB 2; Length 339;
Best Local Similarity 59.2%; Pred. No. 5.5e-78;
Matches 181; Conservative 50; Mismatches 75; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHRLVGSALVRKLEQGFNLVKTAEGLDTRQADVSEFFSQEKPVYVI 62
DB 28 DRAVPIYAGHRLVGSALVRFEASGFTNLVRSRDELDLTDRAATDFVSETRPQVII 87
QY 63 LAAKVGGIHANNTPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYPKFAPQPI 122

```

```

DB 88 DAAARVGGIMANNTPADFLSENLRIQTNLLDAAVAVRPRLFLFGSSCIYPKYAPQPIH 147
QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDAISGMPTNLYGPDNDNPHFPHSV 182
DB 148 ESALLTGPLEPTNDAYAIKAGILQVQAVRRQYGLAWISAMPTNLYGPDNFSGSHL 207
QY 183 LPALMRRPHEAKVNGAEVWVGTSPLREFLHVDDLDADACVFLDDRYSGLEHVNIGSQ 242
DB 208 LPALIRRYEEAKAGGAEEVNTGCTPRRELHVDDLDASACFLLEHFDGPNHVNVTGV 267
QY 243 EYTIRELAEVLKVEVGFEGKLGWCTKPDGTPRKLMDSKSLASLGWTPKVSURDGLSOTY 302
DB 268 DHSISEIADMVATAVGIETRWDPKPDGTPRKLLDVLSALRELGNRPRIALKDGDATV 327
QY 303 DWYLNK 308
DB 328 SWYRTN 333

RESULT 5
ID Q9RCB2 PRELIMINARY; PRT; 321 AA.
AC Q9RCB2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GDP-L-fucose synthetase.
GN FCL.
OS Yersinia pseudotuberculosis (type O:1b).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=109458;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA3606;
RX MEDLINE=20392461; PubMed=10931327;
RA Skurnik M., Peippo A., Ervela E.;
RA "Characterization of the O-antigen gene clusters of Yersinia
RT pseudotuberculosis and the cryptic O-antigen gene cluster of Yersinia
RT pestis shows that the plague bacillus is most closely related to and
RT has evolved from Y. pseudotuberculosis serotype O:1b.";
RL Mol. Microbiol. 37:316-330(2000).
DR EMBL; AJ251712; CAB63301.1; -
DR HSSP; P32055; IBSV.
DR InterPro; IPR00534; Semialdh.dh.
DR Pfam; PF01118; Semialdehyde dh; 1.
SQ SEQUENCE 321 AA; 36289 MW; 10F97CC0D2ABCCB5 CRC64;

Query Match 58.9%; Score 969; DB 2; Length 321;
Best Local Similarity 59.1%; Pred. No. 7.7e-78;
Matches 188; Conservative 48; Mismatches 72; Indels 10; Gaps 2;

QY 3 DKSAKIFVAGHRLVGSALVRKLEQGFNLVKTAEGLDTRQADVSEFFSQEKPVYVI 62
DB 2 DKK-RVFIAGHRLVGSALVRQLENRNDIELIIRTELDLMSQSAVQKFFATEKIDELY 60
QY 63 LAAKVGGIHANNTPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYPKFAPQPI 122
DB 61 LAAKVGGIQANNTPAFIYQNLMECNIIHAHLAQKLLFLGSSCIYPKLAQPM 120
QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDAISGMPTNLYGPDNDNPHFPHSV 182
DB 121 EBALLTGVLEPTNEPYAIKAGIKLCSYNNRQYGRDYSRVNPTNLYGPDNDNPHFPHSV 180
QY 183 LPALMRRPHEAKVNGAEVWVGTSPLREFLHVDDLDADACVFLDDRYSGLEHVNIGSQ 242
DB 181 IPALLRRPHEAKIRNDKENVVWVGTSKPMREFLHVDDMAAASVHVMELSDQIYQTNQPM 240
QY 234 EHVNTSGOEVITRELAEVLKVEVGFEGKLGWCTKPDGTPRKLMDSKSLASLGWTPKVS 293
DB 241 SHINVTGVDCTIRELAETMAKVGVFTGNLVFDSTKPDGTPRKLMDSVSLAKLHCYQIS 300
QY 294 LRDGLSQTVDWYLNKVCN 311

```



```

QY 183 LPALMRPFHEAKVNGAEVVVWGTSPLREFLHVDDADACVFLLDYRYSGLGHEVHNIGSGQ 242
DB 191 LPALIRIERYDEKASGAPNTNMGTTGTPRRELLHVDDLASACVLLGHEFDGTPHVNVTGTGI 250
QY 243 EVTIRELAELVKEVGFEGKLGWDCCTKPDCTPKLMDSSKLASLGWTPKVSRLDGLSQTY 302
DB 251 DHTIGIAEWASAVGYSGETRWDPKPGTGPTRKLLDVSVLRAGWRPSIALRDGIEATV 310
QY 303 DWY 305
DB 311 AWY 313

RESULT 8
Q8XGS8 PRELIMINARY; PRT; 321 AA.
AC Q8XGS8;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Bifunctional GDP fucose synthetase in colanic acid biosynthesis
DE (EC 1.-) (GDP-fucose synthetase) (EC 5.1.1.3.-).
GN WCAG OR STM2108 OR STV2320.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Churcher C., Mungall K.L., Bentley K., Chillingworth T., Connerton P.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque R., Hien T.F., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852 (2001).
RL EMBL; AF008793; AAL21012.1; -
DR EMBL; AL627273; CAD02471.1; -
DR InterPro; IPR001509; Epimerase_Dh.
DR Pfam; PF01370; Epimerase; 1.
KW Oxidoreductase; isomerase; Complete proteome.
SQ SEQUENCE 321 AA; 35815 MW; 2139C7B9C6815D4B CRC64;

Query Match 57.4%; Score 944.5; DB 16; Length 321;
Best Local Similarity 58.5%; Pred. No. 1.2e-75;
Matches 182; Conservative 53; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLGVSIAVRKLOEQGFTNLVLTKEALDLTRQADVSEFSPKPYVILAAA 66
DB 5 RIFVAGHGMVGSIAVRQLAQRGDVLRTRDELDDLDGRAVQAFFAGAGIDQVYLA 64
QY 67 KVGGIHANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPESAL 126
DB 65 KVGGIANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPESAL 124

```

```

QY 127 LTASLEPTNEWAIKIAIKTCQAYRIQHWDAISGMTNLYGPNDFHPNSHVLPA 186
DB 125 LQGTLEPTNEPVAIAKIAIKLCESYNROYGRDYRSMPTNLYGPHDNFHPNSHVLPA 184
QY 187 MRRERHEAKVNGAEVVVWGTSPLREFLHVDDADACVFLLDYRYSGLGHEVHN 237
DB 185 LRRHEAQAOSHAPVVVWGSGTMRREFLHVDDADACVFLLDYRYSGLGHEVHN 244
QY 238 IGSQGVITIRELAELVKEVGFEGKLGWDCCTKPDCTPKLMDSSKLASLGWTPKVSRLD 297
DB 245 VGTGVDCITIRELAQITIAKVVGQGRVVFDAARPDGTPRKLLDVTRLHLQHWHEISLEAG 304
QY 298 LSQTYDWYLN 308
DB 305 LAGTYQWFLEN 315

RESULT 9
Q8X4R4 PRELIMINARY; PRT; 321 AA.
AC Q8X4R4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Putative nucleotide di-P-sugar epimerase or dehydratase (GDP-fucose
DE synthetase chain A).
GN WCAG OR Z3216 OR ECS2857.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller E.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22 (2001).
RL EMBL; AR05431; AAG57112.1; -
DR EMBL; AP02559; BAB36280.1; -
KW Complete proteome.
SQ SEQUENCE 321 AA; 36183 MW; 9511F1DFBD3D8058 CRC64;

Query Match 57.4%; Score 944.5; DB 16; Length 321;
Best Local Similarity 57.9%; Pred. No. 1.2e-75;
Matches 180; Conservative 55; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLGVSIAVRKLOEQGFTNLVLTKEALDLTRQADVSEFSPKPYVILAAA 66
DB 5 RIFVAGHGMVGSIAVRQLAQRGDVLRTRDELDDLDGRAVQAFFAGAGIDQVYLA 64
QY 67 KVGGIHANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPESAL 126
DB 65 KVGGIANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPESAL 124
QY 127 LTASLEPTNEWAIKIAIKTCQAYRIQHWDAISGMTNLYGPNDFHPNSHVLPA 186

```



```
Db 125 LGTLEPTNEPYAIAKIAGIKLCEYNQYGRDYRSVMPNTLYGPHDNFPHNSHVIPAL 184
Qy 187 MRRFHEAKVNGAEVWVGTSPLRFLHVDDLDAC-----VFLLDRYSGLEHVN 237
Db 185 LRRFHEAQAQAPDVWVGSGTMRFLHVDDMAAASIHVMELAEHFWLENTQPMLSHIN 244
Qy 238 IGSGQVTTIRELAELVKEVVGFGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297
Db 245 VGTGVDTCTIRELAQTAKVVGQYGRVDFDASKPDGTPRKLDDVTRLHLQHWHEISLEAG 304
Qy 298 LSQTYDWYLN 308
Db 305 LASTYQWFLN 315
```

## RESULT 10

```
Q9F7A3 PRELIMINARY; PRT; 321 AA.
AC Q9F7A3
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE GDP-L-fucose synthetase.
GN FCL.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=20461159; PubMed=11004393;
RA Stevenson G., Lan R., Reeves P.R.;
RT "The colanic acid gene cluster of salmonella enterica has a complex
RT history."
RL FEMS Microbiol. Lett. 191:11-16(2000).
DR EMBL; AF285084; AAC24814.1; -.
DR HSP; P32055; IBSV.
SQ SEQUENCE 321 AA; 35785 MW; BCB971B11277799F CRC64;
```

```
Query Match 56.4%; Score 928.5; DB 2; Length 321;
Best Local Similarity 57.9%; Pred. No. 3.1e-74;
Matches 180; Conservative 53; Mismatches 69; Indels 9; Gaps 1;
```

```
Qy 7 KIFVAGHGLVGSATVRKLOEQGFTNLVLTAKHAEIDLTRQADVESFSEKPKVYVILAAA 66
Db 5 RIFVAGHGMVGSATVRKLOEQGFTNLVLTAKHAEIDLTRQADVESFSEKPKVYVILAAA 64
Qy 67 KUGGIHANNTPADFGVNLQIOTNVIHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 126
Db 65 KUGGIHANNTPADFGVNLQIOTNVIHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 124
Qy 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHGWDALSGMPTNLYGPNDFPHNSHVLPAL 186
Db 125 LGTLEPTNEPYAIAKIAGIKLCEYNQYGRDYRSVMPNTLYGPHDNFPHNSHVIPAL 184
Qy 187 MRRFHEAKVNGAEVWVGTSPLRFLHVDDLDACVFLDD-----RYSGLEHVN 237
Db 185 LRRFHEAQAQAPDVWVGSGTMRFLHVDDMAAASIHVMELAEHFWLENTQPMLSHIN 244
Qy 238 IGSGQVTTIRELAELVKEVVGFGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297
Db 245 VGTGVDTCTIRELAQTAKVVGQYGRVDFDASKPDGTPRKLDDVTRLHLQHWHEISLEAG 304
Qy 298 LSQTYDWYLN 308
Db 305 LASTYQWFLN 315
```

## RESULT 11

```
Q9S5F8 PRELIMINARY; PRT; 321 AA.
ID, Q9S5F8
```

```
AC Q9S5F8;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE GDP-L-fucose pathway enzyme (Fucose synthetase).
GN WBHF OR FCI OR Z3197 OR EGS2838.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=184;
RX MEDLINE=99240837; PubMed=102222209;
RA Shimizu T., Yamaaki S., Tsukamoto T., Takeda Y.;
RT "Analysis of the genes responsible for the O-antigen synthesis in
RT enterohaemorrhagic Escherichia coli O157."
RL Microb. Pathog. 26:235-247(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamowski K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RN Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AB008676; BAA77731.1; -.
DR EMBL; AE005429; AAG57093.1; -.
DR EMBL; AP002559; BAB36261.1; -.
DR HSP; P32055; IBSV.
KW Complete proteome.
SQ SEQUENCE 321 AA; 36268 MW; 8C9A5311115E33791 CRC64;
```

```
Query Match 55.6%; Score 915.5; DB 16; Length 321;
Best Local Similarity 57.2%; Pred. No. 4.4e-73;
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;
```

```
Qy 7 KIFVAGHGLVGSATVRKLOEQGFTNLVLTAKHAEIDLTRQADVESFSEKPKVYVILAAA 66
Db 5 RIFVAGHGMVGSATVRKLOEQGFTNLVLTAKHAEIDLTRQADVESFSEKPKVYVILAAA 64
Qy 67 KUGGIHANNTPADFGVNLQIOTNVIHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 126
Db 65 KUGGIHANNTPADFGVNLQIOTNVIHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 124
Qy 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHGWDALSGMPTNLYGPNDFPHNSHVLPAL 186
Db 125 LGTLEPTNEPYAIAKIAGIKLCEYNQYGRDYRSVMPNTLYGPHDNFPHNSHVIPAL 184
Qy 187 MRRFHEAKVNGAEVWVGTSPLRFLHVDDLDACVFLDD-----GLEHVN 237
Db 185 LRRFHEAQAQAPDVWVGSGTMRFLHVDDMAAASIHVMELAEHFWLENTQPMLSHIN 244
Qy 238 IGSGQVTTIRELAELVKEVVGFGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297
Db 245 IGTGIDCTICELAETIAKVGKYGKHTFTDTPKDGAPRKLDDVTRLHLQHWNHKTLHGK 304
```

```

QY 298 LSQTYDWYLN 308
DB 305 LENTYNWFLEN 315

RESULT 12
OBS340 PRELIMINARY; PRT; 323 AA.
ID O85340
AC O85340;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Fucose synthetase Fcl.
GN FCL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]_TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN=C664-1992;
RX MEDLINE=98339851; PubMed=9673232;
RA Wang L., Reeves P.R.;
RT "Organization of Escherichia coli O157 O antigen gene cluster and
RT identification of its specific genes.";
RL Infect. Immun. 66:3545-3551 (1998).
DR EMBL; AF061251; AAC32346.1; -.
DR HSSP; P32055; 1BSV.
SQ SEQUENCE 323 AA; 36531 MW; 382A03B17CABD6E9 CRC64;

Query Match 55.6%; Score 915.5; DB 2; Length 323;
Best Local Similarity 57.2%; Pred. No. 4.4e-73;
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;

QY 7 KIFVAGHRLGVSATVRKLEOGFTNLVLKTHAELDLTRQADVESFSPKPVYVILAAA 66
DB 7 RFIAGHQMGVSATIRRLKQDDVELVTRDELMLDSSAVLDFPSSQKIDQVYVILAAA 66

QY 67 KVGGIHANNYPADFTGVNLQIOTNVHSAYEHGVKKLFLGSSCIYKPKFAPQIPESAL 126
DB 67 KVGILANSSYPADFTYENIMIEANVTHAAKNNVKKLFLGSSCIYKPKLAHQPIMEDEL 126

QY 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHGWDASGMPTNLGPNDFHNSHVLPA 186
DB 127 LOGKLEPTNEPYAIKAGIKLCSYNRQGRDYRSVMPNTLYGPNDFHNSHVLPA 186

QY 187 MRPFHAKVNGAEVVVWGTGSPRLREFLHVDDLADACVFLDDR-YS-----GLEHVN 237
DB 187 LARFHDVENVSPNVVWGTGSPKREFLHVDDNASASIVMEWPYDIWQNTKVMLSHIN 246

QY 238 IGSQGEVTIRELAELVKEVVGEGKLGWDCPTKPGTPRKLMDSSKLASLGWTPKVSRLRDG 297
DB 247 IGTGIDICTICELAEITAKVVGKGHITFTDTPDGAPKLLDVTLLHQLGWNHKITLHKG 306

QY 298 LSQTYDWYLN 308
DB 307 LENTYNWFLEN 317

RESULT 13
OBS6L7 PRELIMINARY; PRT; 331 AA.
ID Q8U6L7
AC Q8U6L7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE GDP-fucose synthetase.
GN FCL OR ATU4790 OR AGR L.185.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323 (2001).

RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurolo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009407; AAL45584.1; -.
DR DR EMBL; AE008206; AAK88659.1; -.
KW Complete proteome.
SQ SEQUENCE 331 AA; 36726 MW; 4DB36D98CEE75380 CRC64;

Query Match 54.4%; Score 895.5; DB 16; Length 331;
Best Local Similarity 55.5%; Pred. No. 2.7e-71;
Matches 166; Conservative 59; Mismatches 73; Indels 1; Gaps 1;

QY 7 KIFVAGHRLGVSATVRKLEOGFTNLVLKTHAELDLTRQADVESFSPKPVYVILAAA 66
DB 18 RYVWAGHTGMVGSALVREREN-CEILKVSRELDLTRYETEQWMAARPOVIFVAAA 76

QY 67 KVGGIHANNYPADFTGVNLQIOTNVHSAYEHGVKKLFLGSSCIYKPKFAPQIPESAL 126
DB 77 KVGIIAANAYPADFTLYTTLISMNIMKSAADIGVKLLWMGSSCIYKPKFAAQITENAL 136

QY 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHGWDASGMPTNLGPNDFHNSHVLPA 186
DB 137 LTGPLEPTNEAYAIKAIKALKLSQFYISIQVGLNCVSMPTNIYGLNDFDPOSSHVIPAM 196

QY 187 MRPFHAKVNGAEVVVWGTGSPRLREFLHVDDLADACVFLDDRYSGLHVNIGSGQRTI 246
DB 197 IRRMEHAKISGQNKIVLWGTGSPRLREFLHVDDLADACCFLMKSSAHFPLINIGSGREIS 256

QY 247 RELAEIVKEVVGEGKLGWDCPTKPGTPRKLMDSSKLASLGWTPKVSRLDGLSQTVDWY 305
DB 257 RNLAHLIAGIVGEGQIVFTSKPDGAPKRLDCSRNALGNSTVELRYGIQDLYENW 315

RESULT 14
OBSU14 PRELIMINARY; PRT; 360 AA.
ID Q8VU14
AC Q8VU14;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase.
GN FCL.
OS Bacteroides fragilis.
OC Bacteria; CF group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=817;
RN [1]_TaxID=817;
RP SEQUENCE FROM N.A.
SQ STRAIN=NCTC 9343;

```

RA Coyne M.J., Comstock L.E.;  
RT "A Conserved Region of the Bacteroides fragilis Chromosome Upstream of  
RT the Polysaccharide B Locus Contains Genes Implicated in the Synthesis  
RT of GDP-L-fucose."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF285774; AAU61890.1; -.  
SQ SEQUENCE 360 AA; 40949 MW; FBB327E10A1843B2 CRC64;

Query Match 54.3%; Score 893; DB 2; Length 360;  
Best Local Similarity 51.7%; Pred. No. 5.2e-71;  
Matches 183; Conservative 54; Mismatches 67; Indels 50; Gaps 5;

QY 3 DKSAKIFVAGHRLVGSIAVRKLOEQFTNLVLTAEHLDTLRQADVSEFFSQEKPYYVI 62  
DB 6 DKNKIVVAGHGLVGSIAVRKLOEQFTNLVLTAEHLDTLRQADVSEFFSQEKPYYVI 65  
QY 63 LAAAKVGGIHANNTPADFTGVNLTQNTVIHSAVEHGVKLLFLGSSCIYKFAPOPI 122  
DB 66 LAAAFVGGIMANSIYRADFTYKNTQNTVIHSAVEHGVKLLFLGSSCIYKFAPOPI 125  
QY 123 ESALLTASLEPTNEWVAIAKIGIKTCQAVRIQHGWDALSGMPTNLYGNDNFHPENSHV 182  
DB 126 EDVLLTSPLEYTNEPIAIAKIGIKTCQAVRIQHGWDALSGMPTNLYGNDNFHPENSHV 185  
QY 183 LPALMRFPHEAK-----VNGAE-----EVVYM 204  
DB 186 LPAMIRKVLHAKLKKGDWEAVKDMNLRPVEGISCANSNEETILRLKYGITETETVLM 245  
QY 205 GTGSPRLFLHVDLDADACVFL-----DLY-SGLE-----HVNIGSQEVTIRELAE 252  
DB 246 GTGTPRLFLHVDLDADACVFL-----DLY-SGLE-----HVNIGSQEVTIRELAE 305  
QY 253 VKBVGPEGLKWDCTKPDGTPRLKMDSSKLSLWTPKVSRLDGLSQTYYDYL 306  
DB 306 IYNTVGYQBELTFDSSKPDGTPKRLTDPKSLNHLGWHKIDIEGVQKMYEWL 359

RESULT 15  
Q84611 PRELIMINARY; PRT; 317 AA.  
AC Q84611;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE PBCV-1 fucose synthase.  
GN A295L.  
OS Paraneurium bursaria chlorella virus 1 (PBCV-1).  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.  
OX NCBI\_TaxID=10506;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96187795; PubMed=8614977;  
RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;  
RT "Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map  
RT positions 88 to 182."  
RL Virology 216:102-123(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20013326; PubMed=10544099;  
RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,  
RA Lise A.D., Nickerson K.W., Van Etten J.L.;  
RT "Chlorella virus PBCV-1 encodes a functional homosperrmidine  
RT synthase."  
RL Virology 263:254-262(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20478054; PubMed=11021991;  
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;  
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus  
RT PBCV-1."  
RL Virology 276:27-36(2000).  
RN [4]  
RP SEQUENCE FROM N.A.

RA Van Etten J.L.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Van Etten J.L.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE FROM N.A.  
RA Graves M.V., Van Etten J.L.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP SEQUENCE FROM N.A.  
RA Graves M.V., Van Etten J.L.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP SEQUENCE FROM N.A.  
RA Gurnon J.R., Graves M.V., Van Etten J.L.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U42580; AAC96663.1; -.  
DR HSSP; F32055; 1BSV.  
SQ SEQUENCE 317 AA; 35712 MW; C3A4F99CD9AF82A3 CRC64;

Query Match 53.6%; Score 882; DB 12; Length 317;  
Best Local Similarity 54.6%; Pred. No. 4.1e-70;  
Matches 167; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHRLVGSIAVRKLOEQFTNLVLTAEHLDTLRQADVSEFFSQEKPYYVI 62  
DB 4 DKHSIYVAGHTGMVSSSLVRLLOKEGYNTIITRSSKELDLTNQOEVTAFEMETPEYVF 63  
QY 63 LAAAKVGGIHANNTPADFTGVNLTQNTVIHSAVEHGVKLLFLGSSCIYKFAPOPI 122  
DB 64 LAAAKVGGIHANNTPADFTGVNLTQNTVIHSAVEHGVKLLFLGSSCIYKFAPOPI 123  
QY 123 ESALLTASLEPTNEWVAIAKIGIKTCQAVRIQHGWDALSGMPTNLYGNDNFHPENSHV 182  
DB 124 BEYLMTGFEPTNKPYAERIAIENKDAYKQFCNFCNVNPTNLGNDNDYDLQNGHV 183  
QY 183 LPALMRFPHEAKVNGAEVVMVGTGSPRLFLHVDLDADACVFLDRLDGLSQTYYDYL 242  
DB 184 FVLIRKFEAMINKVPSVCLMGTTGIAARREFLHVDLARGLLIVMEKYNEPGPINIGYS 243  
QY 243 EYVIRELAELVKEVVGPEGLKWDCTKPDGTPRLKMDSSKLSLWTPKVSRLDGLSQTYY 302  
DB 244 DVSISETAELVRDIVGYKAIYDTPMDGTLKKLIDSTKTKSLGWEPKISLIDNKLVCV 303  
QY 303 DWYLNK 308  
DB 304 EDFTKN 309

Search completed: June 3, 2003, 05:04:29  
Job time : 67 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:01:31 ; Search time 35 Seconds  
(without alignments)  
856.970 Million cell updates/sec

Title: US-10-089-014-1  
Perfect score: 1646  
Sequence: 1 MSDKSAKIFVAGHRLVLSA.....SLRDGLSQTYDWLKNVNCNR 312

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1646	100.0	312	2	F96758
2	1478	89.8	328	2	F2H15.12 protein -
3	955	58.0	322	1	probable epiA prot
4	944.5	57.4	321	2	GDP-fucose synthet
5	944.5	57.4	321	2	GDP-fucose synthet
6	944.5	57.4	321	2	GDP-fucose synthet
7	943.5	57.3	321	1	hypothetical 36.1
8	915.5	55.6	321	2	fucose synthetase
9	915.5	55.6	321	2	fucose synthetase
10	895.5	54.4	331	2	hypothetical 34.7K
11	895.5	54.4	331	2	GDP-fucose synthet
12	882	53.6	317	2	hypothetical prote
13	877.5	53.3	309	2	nucleoside-diphosph
14	686	41.7	346	2	probable fucose sy
15	596.5	36.2	246	2	gdp-4-dehydro-D-rh
16	567.5	34.5	314	2	ATN2409
17	565.5	34.4	312	1	S74432
18	547.5	33.3	308	2	T44323
19	505	30.7	310	1	E64525
20	495	30.1	310	2	E71981
21	401	24.4	271	1	S12516
22	365.5	22.2	315	2	T16645
23	345.5	21.0	114	2	A13615
24	270.5	16.4	310	2	E90231
25	266	16.2	309	2	C72368
26	264.5	16.1	310	2	S75771
27	264	16.0	341	2	C69149
28	258	15.7	314	2	C70562
29	253.5	15.4	321	2	G70415

30	253	15.4	307	2	D75143	udp-glucose 4-epim
31	250	15.2	305	2	D64326	UDPglucose 4-epime
32	246	14.9	319	2	D86934	probable sugar-nuc
33	245	14.9	306	1	A71183	probable UDP-gluco
34	243	14.8	334	2	T44339	hypothetical prote
35	240	14.6	334	2	S70889	nucleotide sugar e
36	236	14.3	725	2	H98989	FUSON, Nucleoside
37	231.5	14.1	346	2	C72353	hypothetical prote
38	231	14.0	344	2	D75486	probable UDP-gluco
39	230	14.0	308	2	C84072	UDP-glucose 4-epim
40	227.5	13.8	394	2	A75303	UDP-glucose 4-epim
41	226.5	13.8	343	2	E84113	nucleotide sugar e
42	226	13.7	317	2	F90233	dUDP-Glucose 4,6-d
43	224.5	13.6	321	2	G95920	probable epimerase
44	223.5	13.6	329	2	G82785	dUDP-glucose 4-6-d
45	221.5	13.5	318	2	H71145	probable UDP-gluco

## ALIGNMENTS

### RESULT 1

F96758

hypothetical protein T18K17.8 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C;Accession: F96758

R;Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96758

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-312 <STO>

A;Cross-references: GB:AE005173; NID:G6598858; PIDN:AAF18712.1; GSPDB:GN00141

C;Genetics:

A;Gene: T18K17.8

A;Map position: 1

C;Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 100.0%; Score 1646; DB 2; Length 312;

Best Local Similarity 100.0%; Pred. No. 6e-131;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSDKSAKIFVAGHRLVLSAIVRKLOEQGTNLVLKTHAELDLTRQADVFESFSQKPVY	60
Db	1	MSDKSAKIFVAGHRLVLSAIVRKLOEQGTNLVLKTHAELDLTRQADVFESFSQKPVY	60
QY	61	VILAAKVGIGIHANNTPADFIGVNLQIQTNVIHSAYEHGVKLLFLGSSCIYKFAPOP	120
Db	61	VILAAKVGIGIHANNTPADFIGVNLQIQTNVIHSAYEHGVKLLFLGSSCIYKFAPOP	120
QY	121	IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLGPNDFNPENS	180
Db	121	IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLGPNDFNPENS	180
QY	181	HVLPALMRPFHEAKNGAEVVMWGTGSLRFLHVDDLADACVFLDRYSGLHEHNIGS	240
Db	181	HVLPALMRPFHEAKNGAEVVMWGTGSLRFLHVDDLADACVFLDRYSGLHEHNIGS	240
QY	241	GOEVITRELAELVKEVVGVEGKLGWDCTKPDGTPRKLMDSKSLASLGWTPKVSRLRGLSQ	300
Db	241	GOEVITRELAELVKEVVGVEGKLGWDCTKPDGTPRKLMDSKSLASLGWTPKVSRLRGLSQ	300
QY	301	TYDWLKNVNCNR	312

Db 301 TYDWYLNKNCNR 312  
|||||  
RESULT 2  
B86314  
F2H15.12 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 14-Dec-2001  
C:Accession: B86314  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: B86314  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-328 <STO>  
A:Cross-references: GB:AE005172; NID:99665067; PIDN:AAF97269.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase  
Query Match 89.8%; Score 1478; DB 2; Length 328;  
Best Local Similarity 88.5%; Pred. No. 9e-117;  
Matches 276; Conservative 19; Mismatches 17; Indels 0; Gaps 0;  
Qy 1 MSKSAKIFVAGHRLVGSIAVRKLOQGFNTLVKTHAELDLTROADVSEFFSQKPVY 60  
Db 15 MLEKSAKIFVAGHRLVGSIAVRKLOQGFNTLVKTHAELDLTROADVSEFFSQKPVY 74  
Qy 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYPKFAPOP 120  
Db 75 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYPKFAPOP 134  
Qy 121 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 180  
Db 135 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 194  
Qy 181 HVLPMALRRFHEAKVNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGS 240  
Db 195 HVLPMALRRFHEAKVNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGS 254  
Qy 241 GOBVTIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 300  
Db 255 GVBVTIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 314  
Qy 301 TYDWYLNKNCNR 312  
Db 315 TYEWYLNENVQK 326  
RESULT 3  
C70714  
probable epiA protein - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: C70714  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: C70714

A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-322 <COL>  
A:Cross-references: GB:Z79701; GB:AL123456; NID:g3261635; PIDN:CAB02026.1; PID:g1524259  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: epiA  
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase  
Query Match 58.0%; Score 955; DB 1; Length 322;  
Best Local Similarity 59.7%; Pred. No. 9.6e-73;  
Matches 181; Conservative 44; Mismatches 78; Indels 0; Gaps 0;  
Qy 3 DKSAKIFVAGHRLVGSIAVRKLOQGFNTLVKTHAELDLTROADVSEFFSQKPVY 62  
Db 11 DRAARVYIAGHRLVGSIAVRKLOQGFNTLVKTHAELDLTROADVSEFFSQKPVY 70  
Qy 63 LAAAKVGGIHANNYPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYPKFAPOP 122  
Db 71 DAAARVGGILANDTYPADFLSENQIQVNLDAVAARVPRLLFLGSSCIYPKLAPQPIP 130  
Qy 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 182  
Db 131 ESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 190  
Qy 183 LPALMRRFHEAKVNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGSQ 242  
Db 191 LPALMRRFHEAKVNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGSQ 250  
Qy 243 EYVIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 302  
Db 251 DHTIGEIAEMVASAVGSGTETWDPKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 310  
Qy 303 DWY 305  
Db 311 AWY 313  
RESULT 4  
A10768  
GDP-fucose synthetase (EC 5.1.3.-) [imported] - Salmonella enterica subsp. enterica ser.  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: This species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: A10768  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Party, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero  
A:Reference number: AB0502; PMID:11677608  
A:Accession: A10768  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-321 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD02471.1; PID:g16503337; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY2320  
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase  
C:Keywords: isomerase  
Query Match 57.4%; Score 944.5; DB 2; Length 321;  
Best Local Similarity 58.5%; Pred. No. 7.4e-72;  
Matches 182; Conservative 53; Mismatches 67; Indels 9; Gaps 1;  
Qy 7 KIFVAGHRLVGSIAVRKLOQGFNTLVKTHAELDLTROADVSEFFSQKPVY 66  
Db 5 RIFVAGHRLVGSIAVRKLOQGFNTLVKTHAELDLTROADVSEFFSQKPVY 64  
Qy 67 KVGGIHANNYPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYPKFAPOP 126  
Db 65 KVGIVANNYPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYPKFAPOP 124



A;Gene: yefB  
C;Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 57.3%; Score 943.5; DB 1; Length 321;  
Best Local Similarity 57.6%; Pred. No. 8.9e-72;  
Matches 179; Conservative 56; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHGLVGSIAVRKLEQGFNLVLTAKTHAELDLTRQADVESFFSQEKPVVYLAAA 66  
DB 5 RVFIAGHGMVGSIAIRQLQEQGVDELVRTRDELNLLDSRAVHDFPASERIDQVYLA 64

QY 67 KVGGIHANNYPADFTGVNLQIOTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 126  
DB 65 KVGGIHANNYPADFTGVNLQIOTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 124

QY 127 LTASLEPTNEWYAIKAGIKTCQAVRIQHGWDALSGMPTNLYGPNDFPHNSHVPAL 186  
DB 125 LQGTLEPTNEPYAIKAGIKLCEYNQGRDYRSVMPNLYGPNDFPHNSHVPAL 184

QY 187 MRRFHEAKVNGAEVVMVGTGSPLEFLHVDLADAC-----VFLLDRYSGLEHVN 237  
DB 185 LRRFHDVENSPPVVMVGTGSPLEFLHVDLADAC-----VFLLDRYSGLEHVN 237

QY 238 IGSQGVETIRELAELVKEVVGFEKGLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297  
DB 245 VGTGVDCTIRELAETIAKVGVYGRVVDASKPDGTPRKLMDSSKLASLGWTPKVSRLDG 304

QY 298 LSQTYDYLKLN 308  
DB 305 LENTYNWFLEN 315

RESULT 8  
A85829  
fucose synthetase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Dec-2001  
C;Accession: A85829  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimlant, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; PMID:21074935; PMID:11206551  
A;Accession: A85829  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-321 <STO>  
A;Cross-references: GB:AB005174; NID:G12516221; PIDN:AA057093.1; GSPDB:GN00145; UMGF:231  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: fci  
C;Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 55.6%; Score 915.5; DB 2; Length 321;  
Best Local Similarity 57.2%; Pred. No. 2e-69;  
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;

QY 7 KIFVAGHGLVGSIAVRKLEQGFNLVLTAKTHAELDLTRQADVESFFSQEKPVVYLAAA 66  
DB 5 RVFIAGHGMVGSIAIRQLQEQGVDELVRTRDELNLLDSRAVHDFPASERIDQVYLA 64

QY 67 KVGGIHANNYPADFTGVNLQIOTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 126  
DB 65 KVGGIHANNYPADFTGVNLQIOTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 124

QY 127 LTASLEPTNEWYAIKAGIKTCQAVRIQHGWDALSGMPTNLYGPNDFPHNSHVPAL 186  
DB 125 LQGTLEPTNEPYAIKAGIKLCEYNQGRDYRSVMPNLYGPNDFPHNSHVPAL 184

QY 187 MRRFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLDR-YS-----GLEHVN 237  
DB 185 LRRFHDVENSPPVVMVGTGSPLEFLHVDLADACVFLDR-YS-----GLEHVN 237

QY 298 LSQTYDYLKLN 308  
DB 305 LENTYNWFLEN 315

RESULT 10  
A98142  
hypotheical 34.7K protein y4af [imported] - Agrobacterium tumefaciens (strain C58, Cer  
hypotheical 34.7K protein y4af [imported] - Agrobacterium tumefaciens (strain C58, Cer  
C;Species: Agrobacterium tumefaciens  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C;Accession: A98142  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; PMID:11743194  
A;Accession: A98142  
A;Status: preliminary  
A;Molecule type: DNA

QY 238 IGSQGVETIRELAELVKEVVGFEKGLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297  
DB 245 VGTGVDCTIRELAETIAKVGVYGRVVDASKPDGTPRKLMDSSKLASLGWTPKVSRLDG 304

QY 298 LSQTYDYLKLN 308  
DB 305 LENTYNWFLEN 315

RESULT 9  
F90983  
fucose synthetase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 14-Dec-2001  
C;Accession: F90983  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G  
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A;Reference number: A99629; PMID:21156231; PMID:11258796  
A;Accession: F90983  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-321 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA036261.1; PID:G1362306; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: EC82838  
C;Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 55.6%; Score 915.5; DB 2; Length 321;  
Best Local Similarity 57.2%; Pred. No. 2e-69;  
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;

QY 7 KIFVAGHGLVGSIAVRKLEQGFNLVLTAKTHAELDLTRQADVESFFSQEKPVVYLAAA 66  
DB 5 RVFIAGHGMVGSIAIRQLQEQGVDELVRTRDELNLLDSRAVHDFPASERIDQVYLA 64

QY 67 KVGGIHANNYPADFTGVNLQIOTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 126  
DB 65 KVGGIHANNYPADFTGVNLQIOTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 124

QY 127 LTASLEPTNEWYAIKAGIKTCQAVRIQHGWDALSGMPTNLYGPNDFPHNSHVPAL 186  
DB 125 LQGTLEPTNEPYAIKAGIKLCEYNQGRDYRSVMPNLYGPNDFPHNSHVPAL 184

QY 187 MRRFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLDR-YS-----GLEHVN 237  
DB 185 LRRFHDVENSPPVVMVGTGSPLEFLHVDLADACVFLDR-YS-----GLEHVN 237

QY 238 IGSQGVETIRELAELVKEVVGFEKGLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297  
DB 245 VGTGVDCTIRELAETIAKVGVYGRVVDASKPDGTPRKLMDSSKLASLGWTPKVSRLDG 304

QY 298 LSQTYDYLKLN 308  
DB 305 LENTYNWFLEN 315

RESULT 10  
A98142  
hypotheical 34.7K protein y4af [imported] - Agrobacterium tumefaciens (strain C58, Cer  
hypotheical 34.7K protein y4af [imported] - Agrobacterium tumefaciens (strain C58, Cer  
C;Species: Agrobacterium tumefaciens  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C;Accession: A98142  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; PMID:11743194  
A;Accession: A98142  
A;Status: preliminary  
A;Molecule type: DNA



A:Residues: 1-331 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK88659.1; PID:g15158384; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L185  
A:Map position: linear chromosome  
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 54.4%; Score 895.5; DB 2; Length 331;  
Best Local Similarity 55.5%; Pred. No. 1e-67;  
Matches 166; Conservative 59; Mismatches 73; Indels 1; Gaps 1;

QY 7 KIFVAGHRLVGSIAVRKLEQGFNLVLTAEHLDLTRQADVESFFSQEKPVYVILAAA 66  
DB 18 RVWVAGHTGVGSALVRRLEREN-CEILKVSRLDLTRQYETEQQWMAARQVIVFAAA 76  
QY 67 KVGGIHANNYPADFGVNLQIOTNVIHSAYEHGVKKLFLGSSCIYPKFAPQIPESAL 126  
DB 77 KVGIIAANAAYPADFLYNTLISNMIMKSAADIGVEKLLWMGSSCIYPKFAAQITENAL 136  
QY 127 LTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDASLGMTNLYGPNDFPHNSHVLPA 186  
DB 137 LTGPLEPTNEWAIAKIALKLSQFSIQVGLNCVSMPTNIYGLNDFDQSSHVIPAM 196  
QY 187 MRPFHAKVNGAEVVVWGTSPLREFLHVDDLDACVFLDRYSGLFHVNTGSGQEV 246  
DB 197 IRRMHEAKISGQNKIVLWGTGSPRLREFLHVDDLDACVFLDRYSGLFHVNTGSGREI 256  
QY 247 RELAEVLKVEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLDGLSQT 305  
DB 257 RNLHLIAGIVGEGQIVFTSKPDGAPRKLDCSLRNALGWNSTVELYIGIQLYEW 315

RESULT 11  
AB3146  
GDP-fucose synthetase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AB3146  
R:Wood, D.W.; Stetubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AB3146  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAK45584.1; PID:g17743301; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: fcl  
A:Map position: linear chromosome  
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 54.4%; Score 895.5; DB 2; Length 331;  
Best Local Similarity 55.5%; Pred. No. 1e-67;  
Matches 166; Conservative 59; Mismatches 73; Indels 1; Gaps 1;

QY 7 KIFVAGHRLVGSIAVRKLEQGFNLVLTAEHLDLTRQADVESFFSQEKPVYVILAAA 66  
DB 18 RVWVAGHTGVGSALVRRLEREN-CEILKVSRLDLTRQYETEQQWMAARQVIVFAAA 76  
QY 67 KVGGIHANNYPADFGVNLQIOTNVIHSAYEHGVKKLFLGSSCIYPKFAPQIPESAL 126  
DB 77 KVGIIAANAAYPADFLYNTLISNMIMKSAADIGVEKLLWMGSSCIYPKFAAQITENAL 136  
QY 127 LTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDASLGMTNLYGPNDFPHNSHVLPA 186  
DB 137 LTGPLEPTNEWAIAKIALKLSQFSIQVGLNCVSMPTNIYGLNDFDQSSHVIPAM 196

QY 187 MRPFHAKVNGAEVVVWGTSPLREFLHVDDLDACVFLDRYSGLFHVNTGSGQEV 246  
DB 197 IRRMHEAKISGQNKIVLWGTGSPRLREFLHVDDLDACVFLDRYSGLFHVNTGSGREI 256  
QY 247 RELAEVLKVEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLDGLSQT 305  
DB 257 RNLHLIAGIVGEGQIVFTSKPDGAPRKLDCSLRNALGWNSTVELYIGIQLYEW 315

# RESULT 12

TI1792  
hypothetical protein A295L - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: TI1792  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: TI1792  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-317 <GRA>  
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96663.1  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A:Note: A295L  
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 53.6%; Score 882; DB 2; Length 317;  
Best Local Similarity 54.6%; Pred. No. 1.3e-66;  
Matches 167; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHRLVGSIAVRKLEQGFNLVLTAEHLDLTRQADVESFFSQEKPVYV 62  
DB 4 DKHSKIYVAGTGWGSSLVRLQKEGTYNIITRSKELDLINQOEVTAFEMETEYVF 63  
QY 63 LAAAKVGGIHANNYPADFGVNLQIOTNVIHSAYEHGVKKLFLGSSCIYPKFAPQIP 122  
DB 64 LAAAKVGGIHANNYPADFGVNLQIOTNVIHSAYEHGVKKLFLGSSCIYPKESNP 123  
QY 123 ESALLTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDASLGMTNLYGPNDFPHNSH 182  
DB 124 EYLMTGFLPEPTNKPKVAIERIAGIEMCDAYRQFCNFSVMPNTLSGPNDRYDLQ 183  
QY 183 LPALMRPFHAKVNGAEVVVWGTSPLREFLHVDDLDACVFLDRYSGLFHVNTGSG 242  
DB 184 FPLVIRKFEYAMINKVPSVKLWGTGAREFLHVDDLDARGLLIIVMEKYNPEPIN 243  
QY 243 EYVIRELAEVLKVEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLD 302  
DB 244 DVSISELAEVLKVEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLD 303  
QY 303 DWYLNK 308  
DB 304 EDFTKN 309

# RESULT 13

F97168  
nucleoside-diphosphate-sugar epimerase [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 31-Dec-2001  
C:Accession: F97168  
R:Nolling, J.; Bennett, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: F97168  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-309 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK80137.1; PID:g15025174; GSPDB:GN00168

A; Experimental source: Clostridium acetobutylicum ATCC824

C; Genetics:

A; Gene: CAC2179

C; Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 53.3%; Score 877.5; DB 2; Length 309;

Best Local Similarity 53.6%; Pred. No. 3.1e-66;

Matches 164; Conservative 66; Mismatches 75; Indels 1; Gaps 1;

QY 4 KSAKIFVAGHRLVGSALVRKLOEGFTNLVLKTHAELDLTROADVSPFSQEKPVVIL 63

DB 3 EDSKIYIAGHTGVGSAILNLRERRGYKXNVVRTHKELDLMHQESVKKFLEEKPDYVVL 62

QY 64 AAQVGGIHNANTYPADFIGVNLQIOTNVTHSAVEHGVKKLFLGSSCIYKPAPOPIPE 123

DB 63 SAAKVGGIQANISNPVDFLMDNLIIENYVKNISFEVGIENLLFLGSSCIYKPAPOPLKE 122

QY 124 SALLTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLYGPNDFHPENSHVL 183

DB 123 EYLLSGYLEPTNEGYAIKISGLKMCYYSKQGLNYISAMPNSLYGMRDNFPLKTSVM 182

QY 184 PALMRPFHAKVNGAEVVMVGTGSPLEFLHVDDLDACVFLLDYRSGLE-HVNIGSQ 242

DB 183 AALIRPFHAKVSGSGEISIWGSGEQRYRFTYIEDLADGIIIFUMHEGKVGFLNTGCGK 242

QY 243 EVTIRELAELVKEWGFEGKLGWDCPKDGTPRKMLDSSKLASLGWTPKVSRLDGLSQT 302

DB 243 DIKIKLAYKIDVGVFGKNIIFDKSKPDGMFRKMDVSKNSLGHVHYKVELDEGITKY 302

QY 303 DWYLNK 308

DB 303 RMYLNN 308

RESULT 14

probable fucose synthetase Cj1428c [imported] - Campylobacter jejuni (strain NCTC 11168)

C; Species: Campylobacter jejuni

C; Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C; Accession: D81288

R; Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000

A; Title: The genome sequence of the food-borne pathogen Campylobacter jejuni. reveals hyp

A; Reference number: A81250; MUID:20150912; PMID:10688204

A; Accession: D81288

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-346 <PAR>

A; Cross-references: GB:AL139078; GB:AL111168; NID:G6968723; PIDN:CAB73852.1; PID:G696885

A; Experimental source: serotype O2, strain NCTC 11168

C; Genetics:

A; Gene: fcl; Cj1428c

Query Match 41.7%; Score 686; DB 2; Length 346;

Best Local Similarity 42.6%; Pred. No. 4.8e-50;

Matches 145; Conservative 53; Mismatches 102; Indels 40; Gaps 4;

QY 5 SAKIFVAGHRLVGSALVRKLOEGFTNLVLKTHAELDLTROADVSPFSQEKPVVIL 64

DB 4 NSKIYIAGHKGTAGTALVENLQKGFNNLVLTQRLQDLVNQQAQVAKFFKEEKPEYVLT 63

QY 65 AAQVGGIHNANTYPADFIGVNLQIOTNVTHSAVEHGVKKLFLGSSCIYKPAPOPIPE 124

DB 64 AVLPGA-ANVAQRAADFIYENLMQNVIHNSFLNNVKKLVFGSGYMPENAKNPLKEE 122

QY 125 ALLTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLYGPNDFHPENSHVL 184

DB 123 YLFGQDLEYGVSFGAAGIAGIMCESYNIQYGTNFIITLVNLYGTKANFDFGKSRVLP 182

QY 185 ALMRPFH-----EAKVNGAE-----VVWGTGSPLEFLH 215

DB 183 ALLRKPHLAKLISEGNITQIQLDKNNFEEAKYLLHNFISKKSVEIWTGKVRREFT 242

QY 216 VDDLADACVFL-----DRYSGLEHVNIGSQEVTIRELAELVKEVVGFEKLGW 265

DB 243 SDDLADVAIYTMQNDPKDLIKDRKSKNTHINIGTIDYSIKEVALMVKNIVGSGELVF 302

QY 266 DCTKPDGTPRKLMDSSKLASLGWTPKVSRLDGLSQTQYDWY 305

DB 303 NTSRDPSTMDRLMDCSKIHSLGKHKIELKDGIKMWYEWY 342

RESULT 15

gdp-4-dehydro-D-rhamnose reductase (EC 1.1.1.187) [imported] - Brucella melitensis (str)

AH3615

C; Species: Brucella melitensis

C; Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002

C; Accession: AH3615

R; DelVecchio, V.G.; Kaparal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,

; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A; Title: The genome sequence of the facultative intracellular pathogen Brucella meliten

A; Reference number: AD3252; PMID:11756688

A; Accession: AH3615

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-246 <KUR>

A; Cross-references: GB:AE008918; PIDN:AAL54091.1; PID:GI7985049; GSPDB:GN00191

A; Experimental source: strain 16M

C; Genetics:

A; Gene: BME110849

A; Map position: II

C; Keywords: Oxidoreductase

Query Match 36.2%; Score 596.5; DB 2; Length 246;

Best Local Similarity 53.2%; Pred. No. 1e-42; 50; Indels 21; Gaps 4;

Matches 124; Conservative 38; Mismatches 38; Indels 21; Gaps 4;

QY 7 KIFVAGHRLVGSALVRKLOEGFTNLVLKTHAELDLTROADVSPFSQEKPVVILAAA 66

DB 19 KIFVAGHTGVGSAILRLQHEH-CDIITAHSVLDTLQGPTEFISGHRPDVILIAAA 77

QY 67 KVGGIHNANTYPADFIGVNLQIOTNVTHSAVEHGVKKLFLGSSCIYKPAPOPIPE 126

DB 78 RVGGILANSRFPADFLYNNLAIGNLIHAAHQIGVERLLWLGSSCIYPRDAAQPLTEDAL 137

QY 127 LTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLYGPNDFHPENSHVL 186

DB 138 LTGPLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLYGPNDFHPENSHVL 197

QY 187 MRRPFHAKVNGAEV-----VWGTGSPLEFLHVDLADACVFL 227

DB 198 IRRVHEARMRGAEEVCFGAAASPCGNSCMW-TTSPMRACI-----SCAFIM 242

Search completed: June 3, 2003, 05:09:20

Job time : 36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:02:36 ; Search time 21 Seconds  
(without alignments)  
437.141 Million cell updates/sec

Title: US-10-089-014-1  
Perfect score: 1645  
Sequence: 1 MDKSAKIFVAGHGLVSA.....SLRDGLSQTYDWLKNVNR 312

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	463.5	28.2	321	2 US-08-937-972-3	Sequence 3, Appli
2	402	24.4	271	2 US-08-937-972-6	Sequence 6, Appli
3	206.5	12.5	3782	4 US-09-105-537-4	Sequence 4, Appli
4	203.5	12.4	329	4 US-09-036-987A-26	Sequence 26, Appl
5	203.5	12.4	329	4 US-09-370-700-26	Sequence 26, Appl
6	203.5	12.4	337	3 US-09-320-878-15	Sequence 15, Appl
7	199.5	12.1	337	4 US-09-105-537-14	Sequence 14, Appl
8	190.5	11.6	325	4 US-09-194-905-11	Sequence 11, Appl
9	180	10.9	278	4 US-09-149-476-701	Sequence 701, App
10	169.5	10.3	348	4 US-09-113-536-2	Sequence 2, Appli
11	169.5	10.3	348	4 US-09-216-183-2	Sequence 2, Appli
12	169.5	10.3	348	5 PCT-US95-05785-2	Sequence 2, Appli
13	151.5	9.2	306	4 US-09-392-772-6	Sequence 6, Appli
14	151.5	9.2	344	4 US-09-392-772-2	Sequence 2, Appli
15	139	8.4	427	4 US-09-392-772-4	Sequence 4, Appli
16	129.5	7.9	322	2 US-08-576-626A-33	Sequence 33, Appl
17	129.5	7.9	2544	2 US-08-576-626A-32	Sequence 32, Appl
18	129	7.8	327	3 US-09-154-874-8	Sequence 8, Appli
19	125	7.6	366	4 US-09-638-715-2	Sequence 2, Appli
20	125	7.6	366	4 US-09-638-715-4	Sequence 4, Appli
21	124	7.5	337	2 US-08-923-856-4	Sequence 4, Appli
22	124	7.5	337	3 US-09-216-294-4	Sequence 4, Appli
23	122.5	7.4	326	3 US-09-154-874-9	Sequence 9, Appli
24	107	6.5	153	4 US-09-392-772-8	Sequence 8, Appli
25	106	6.4	299	2 US-08-576-626A-35	Sequence 35, Appl
26	104	6.3	338	3 US-08-722-184-8	Sequence 8, Appli
27	104	6.3	338	4 US-09-043-937A-12	Sequence 12, Appli

28	100	6.1	335	3 US-08-722-184-2	Sequence 2, Appli
29	100	6.1	335	4 US-09-043-937A-6	Sequence 6, Appli
30	96.5	5.9	290	2 US-08-576-626A-34	Sequence 34, Appl
31	95	5.8	335	3 US-08-722-184-10	Sequence 10, Appl
32	95	5.8	335	4 US-09-043-937A-14	Sequence 14, Appl
33	95	5.8	1722	4 US-09-194-612A-1	Sequence 1, Appli
34	94.5	5.7	342	3 US-08-722-184-6	Sequence 6, Appli
35	94.5	5.7	342	4 US-09-043-937A-10	Sequence 10, Appl
36	90.5	5.5	338	3 US-08-722-184-4	Sequence 4, Appli
37	90.5	5.5	338	4 US-09-043-937A-8	Sequence 8, Appli
38	90	5.5	320	4 US-09-134-001C-4439	Sequence 4439, Ap
39	89.5	5.4	371	4 US-09-043-937A-4	Sequence 4, Appli
40	89	5.4	181	4 US-09-194-905-6	Sequence 6, Appli
41	89	5.4	924	4 US-09-267-311-2	Sequence 2, Appli
42	88.5	5.4	1447	4 US-09-041-886-25	Sequence 25, Appl
43	88.5	5.4	1447	5 PCT-US94-05277-2	Sequence 2, Appli
44	86	5.2	312	4 US-09-475-316A-108	Sequence 108, App
45	85	5.2	320	2 US-08-846-762-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1  
US-08-937-972-3  
; Sequence 3, Application US/08937972  
; Patent No. 5932443  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: ANTIGENS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/937,972  
; FILING DATE: Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0400 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 321 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADNOT04  
; CLONE: 1318190  
US-08-937-972-3

Query Match 28.2%; Score 463.5; DB 2; Length 321;

Best Local Similarity 35.4%; Pred. No. 4.8e-42;  
Matches 111; Conservative 64; Mismatches 124; Indels 15; Gaps 6;

QY 5 SAKIFVAGHRLGSAIVRKLQEGF-----TNLVKTHAELDLTROADVESFSSQSKPVY 60  
DB 7 SMRIIVTGGGLGVKAI-QKVADGAGLPGEWVSVSSKDADLTDTAQTRALFEKVQPTH 65  
QY 61 VILAAKVGSIHANNTYPADFIGVNLQIOTNVIHSAHEGVKLLFLGSSCIYPKPAQP 120  
DB 66 VIHLAAWVGGLFRNIKNLDFWRKVNHDNVLHSAFEVGARKVSVCLSTCIPDKTYP 125  
QY 121 IPSALLTASLEPTNEWAIAGIKTCQAYRIQHWDAISGMPTNLYGPNNDHPENS 180  
DB 126 IDETMIHNGPPHNSFGYSYAKRMIDVQNRAYFQOYGCCTTAVIPTNVEGPDNFNIEDG 185  
QY 181 HVLPAALRRFHEAKVGAEEVVMGTGSPLEFLHVDLADACVFLLDYSGLEHV--NI 238  
DB 186 HVLPLGLIHKVHLAKSGS-ALTVMGNGNPRRQFIYSLDLAQLFIWVLRNEYVEPIILSV 244  
QY 239 GSGQEVTTIRELAELVKEVGFEGKLGWDCCTKPGCTPRKLMDSKSLAS----LQWTPKVSL 294  
DB 245 GEEDEVSIEKAAEAARVAEMDFHGEVTFDTKSDGQFKTASNSKLTYPDFRFTP--P 301  
QY 295 RDGLSOTYDWYLN 308  
DB 302 QKAVKETCAWFTDN 315

## RESULT 2

US-08-937-972-6  
; Sequence 6, Application US/08937972  
; Patent No. 5932443  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: ANTIGENS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/937,972  
; FILING DATE: Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0400 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 271 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:

; LIBRARY: GenBank  
; CLONE: 199586  
; US-08-937-972-6  
Query Match 24.4%; Score 402; DB 2; Length 271;  
Best Local Similarity 37.3%; Pred. No. 1.8e-35;  
Matches 98; Conservative 47; Mismatches 110; Indels 8; Gaps 4;

QY 5 SAKIFVAGHRLGSAIVRKLQEGF-----TNLVKTHAELDLTROADVESFSSQSKPVY 60  
DB 7 SMRIIVTGGGLGVKAI-QKVADGAGLPGEWVSVSSKDADLTDTAQTRALFEKVQPTH 65  
QY 61 VILAAKVGSIHANNTYPADFIGVNLQIOTNVIHSAHEGVKLLFLGSSCIYPKPAQP 120  
DB 66 VIHLAAWVGGLFRNIKNLDFWRKVNHDNVLHSAFEVGARKVSVCLSTCIPDKTYP 125  
QY 121 IPSALLTASLEPTNEWAIAGIKTCQAYRIQHWDAISGMPTNLYGPNNDHPENS 180  
DB 126 IDETMIHNGPPHNSFGYSYAKRMIDVQNRAYFQOYGCCTTAVIPTNVEGPDNFNIEDG 185  
QY 181 HVLPAALRRFHEAKVGAEEVVMGTGSPLEFLHVDLADACVFLLDYSGLEHV--NI 238  
DB 186 HVLPLGLIHKVHLAKSGS-ALTVMGNGNPRRQFIYSLDLAQLFIWVLRNEYVEPIILSV 244  
QY 239 GSGQEVTTIRELAELVKEVGFEG 261  
DB 245 GEEDEVSIEKAAEAARVAEMDFNG 267

## RESULT 3

US-09-105-537-4  
; Sequence 4, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT FILING DATE: 1998-06-26  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 3782  
; TYPE: PR1  
; ORGANISM: Streptomyces venezuelae  
; US-09-105-537-4  
Query Match 12.5%; Score 206.5; DB 4; Length 3782;  
Best Local Similarity 24.2%; Pred. No. 2.2e-12;  
Matches 84; Conservative 52; Mismatches 140; Indels 71; Gaps 11;

QY 3 DKSAKIFVAGHRLGSAIVRKLQEGFNTNLVLTTHAELD-LT----- 44  
DB 2251 DQAMELLVTGGAGFIGSHFVRQLLAGAYDPDPADEVIVLDSLTAGNANLANLAPVDADPRL 2310  
QY 45 -----ROADVESFSSQSKPVYVILAAKVGSIHANNTYPADFIGVNLQ 87  
DB 2311 RFVHGDIRDAGLLARELGRGVDALVHFAAEASHVDRSAGASV-----FTETNVQ 2358  
QY 88 IOTNVIHSAHEGVKLLFLGSSCIYPKPAQPPIESALLTASLPTNEWAIAGIK 147  
DB 2359 GTQTLQCAVDAGVGRVHVSTDEVIGSIDGSWTES-----SPLEP-NSPYAASKAGSDL 2413  
QY 148 TCQAYRIQHWDAISGMPTNLYGPNNDHPENSHVLPALMRFRHBAKVNGAEVVMGTG 207  
DB 2414 VARAVHRTYGLDVIRTRCCNNVGPYQ--HPEK-----LIPLVFTNLLDGG-TLPLYGDG 2464  
QY 208 SPLREFLHVDLADACVFLLDYSGLEHVNTGSGQEVTTIRELAELVKEVGFEGKLGWDC 267  
DB 2465 ANVREWHVHTDDHCRGIALVLGAGRAGEITYHGGGLNRELTLGILLDSLGAD----WSS 2520



```

; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27 OF 09/141,908
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 15
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-15

Query Match      12.4%; Score 203.5; DB 3; Length 337;
Best Local Similarity 24.5%; Pred. No. 1e-13;
Matches 84; Conservative 49; Mismatches 139; Indels 71; Gaps 11;

QY 7 KIFVAGHRLVGSIAVRKLEQGFNTLVKTHAELD-LT-----FTTNVGGTQT 109
Db 2 RLLVTGGAGFIGSHFVRQLLAGAYPDVPADEVILDSLTVYAGNRANLAPVDADPRLRFVH 61
QY 45 -----ROADVESFSEKPYVIVLAAGKVGIIHANNTPYADFIGVNLQIQTN 91
Db 62 GDIRDAGLLARELGRGVDIVHFAAESHVDRSIAGASV-----FTTNVGGTQT 109
QY 92 VIHSAYEHGVKLLFLGSSCIYKPAQPIPESSALLTASLEPTNEWYAIKAGIKTCA 151
Db 110 LLOCAVDAGVRVHVSTDEVYSGDSWTES-----SPLEP-NSPYAASKAGSGLVARA 164
QY 152 YRIQHWDAISGMPTNLYGPNDFHPENSHVLPALMRPFHEAKVNGAEVVMVGTGSP 211
Db 165 YHRTYGLDVRITRCCNNYGPYQ--HPEK-----LIPLFVTNLLDGG-TLPLYGDCANVR 215
QY 212 EFLHVDLADACVFLLDYRSLGHEHVNIGSQEVTIRELAELVKEVVGFEKLGWDC 271
Db 216 EWHTDDHCRGIALVLAGGRAGEIYHGGLELTNRLELTGILLDSLGAD-----WSSVRKV 271
QY 272 GTP-----RKLMDSSKL-ASLGWTPKVSRLDGLSOTYDWYLNK 308
Db 272 ADKRGHDLRYSLDGGKIERELGRYPQVSPADGLARTVRYREN 314

; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
; TITLE OF INVENTION: GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESS: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,905
; FILING DATE: 29-JUL-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP97/02826
; FILING DATE: 30-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19622783.6
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 026083/0193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids

```

```

US-09-105-537-14
Query Match      12.1%; Score 199.5; DB 4; Length 337;
Best Local Similarity 24.2%; Pred. No. 2.8e-13;
Matches 83; Conservative 50; Mismatches 139; Indels 71; Gaps 11;

QY 7 KIFVAGHRLVGSIAVRKLEQGFNTLVKTHAELD-LT-----FTTNVGGTQT 44
Db 2 RLLVTGGAGFIGSHFVRQLLAGAYPDVPADEVILDSLTVYAGNRANLAPVDADPRLRFVH 61
QY 45 -----ROADVESFSEKPYVIVLAAGKVGIIHANNTPYADFIGVNLQIQTN 91
Db 62 GDIRDAGLLARELGRGVDIVHFAAESHVDRSIAGASV-----FTTNVGGTQT 109
QY 92 VIHSAYEHGVKLLFLGSSCIYKPAQPIPESSALLTASLEPTNEWYAIKAGIKTCA 151
Db 110 LLOCAVDAGVRVHVSTDEVYSGDSWTES-----SPLEP-NSPYAASKAGSGLVARA 164
QY 152 YRIQHWDAISGMPTNLYGPNDFHPENSHVLPALMRPFHEAKVNGAEVVMVGTGSP 211
Db 165 YHRTYGLDVRITRCCNNYGPYQ--HPEK-----LIPLFVTNLLDGG-TLPLYGDCANVR 215
QY 212 EFLHVDLADACVFLLDYRSLGHEHVNIGSQEVTIRELAELVKEVVGFEKLGWDC 271
Db 216 EWHTDDHCRGIALVLAGGRAGEIYHGGLELTNRLELTGILLDSLGAD-----WSSVRKV 271
QY 272 GTP-----RKLMDSSKL-ASLGWTPKVSRLDGLSOTYDWYLNK 308
Db 272 ADKRGHDLRYSLDGGKIERELGRYPQVSPADGLARTVRYREN 314

RESULT 8
US-09-194-905-11
; Sequence 11, Application US/09194905
; Patent No. 6306627
; GENERAL INFORMATION:
; APPLICANT: DECKER, Heinrich
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
; TITLE OF INVENTION: GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESS: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,905
; FILING DATE: 29-JUL-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP97/02826
; FILING DATE: 30-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19622783.6
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 026083/0193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids

```

```
;
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-194-905-11

Query Match      11.6%; Score 190.5; DB 4; Length 325;
Best Local Similarity 25.4%; Pred. No. 2.6e-12;
Matches 85; Conservative 50; Mismatches 136; Indels 63; Gaps 14;

QY 8 IFVAGHRLGVGAIVKRLQEQG-----FTNLVLKTHA-----ELDL 43
Db 5 ILVTGGAGIRAYVRLLSPAGPGVAVTLDKLTYAGSLARLHRAVRDHPGLTFVQGV 64
QY 44 TRQADVSPFSQKPYVYILAAKVGGIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKK 103
Db 65 CDTALVDTLAARHDDIVFAEAESHV---DRSITDSGAFRTNVLGTQVLLDAALRHGVRT 121
QY 104 LFLGSSCIYKFAFQPIESPESALLTAS-LEPTNEWYAIKAGIKTCQAYRIQHGWDALS 162
Db 122 FHVSTDEVYGS-----LPHGAASDPLLTPTSP-YAASKAASDLMALAHRTHTGLDVRV 175
QY 163 GMPNTLYGNDFNPENSHVLPALMRPFHEAKVNGAEVVMWGTGSPLEFLHVDLADA 222
Db 176 TRCSNNFGPHQ--HPEK-----LIPRFLTSLSGG-TVPLYGDGHVDRDLHVDHVR 226
QY 223 CVFLDRYSGL--EHNIGSGQEVITRELAELV-----KEVVGFEKGLGMDCTKPDG 272
Db 227 VE--LVRVSGRPGCEIYNIIGGTSPLNLELTHRLALCGAGPERIVHVENRKGH----- 278
QY 273 TPKLMDSKL-ASLGWTPKVSIRDLGSLQTYDWY 305
Db 279 -RRYAVDHSKITAEIYGLRPRDTFATALADTAKWY 311

RESULT 9
US-09-149-476-701
; Sequence 701, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
```





```
/ FILING DATE: 11 MAY 1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: FERRARO, GREGORY D.
/ REGISTRATION NUMBER: 36,134
/ REFERENCE/DOCKET NUMBER: 325800-430
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 201-994-1700
/ TELEFAX: 201-994-1744
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 348 AMINO ACIDS
/ TYPE: AMINO ACID
/ STRANDEDNESS:
/ TOPOLOGY: LINEAR
/ MOLECULE TYPE: PROTEIN
/ US-09-113-536-2

Query Match 10.3%; Score 169.5; DB 4; Length 348;
Best Local Similarity 21.9%; Pred. No. 5.5e-10;
Matches 80; Conservative 58; Mismatches 135; Indels 93; Gaps 15;

QY 7 KIFVAGHRLGVSIAVRKLOEQGFTNLVLT-----HA 39
Db 4 KVLVTGAGYIGSHTVLELEAGYLPVVDNFHNAFRGGSLPESLRRVQELTGRSVEFE 63
QY 40 ELDLTQADVESFSEKFPVYVILAAKVGGIHA---NNTYPADFIGVNLQIOTNVHSA 96
Db 64 EMDILDQAGLQRLFKK---YSFMAVIHFAGLKAVGESVQKPLDYRVNLTGTIQLLEIM 119
QY 97 YEHGVKKLLPLGSSCIYPKPAQPIPEASALLTASLEP--TNEWATAKIAGIKTCQAYRI 154
Db 120 KAHGVKNLVFSSSATVYGNPQYLPLDEAHPGTGCTNPGYKSKFFIEMIRDL--COADKT 177
QY 155 QHGWDAI-----SGMPTNLYGNPNDNFPHNSHVLPAALMRFPHEAK 194
Db 178 ---WNAVLLRYFNPTGAHSGCIGEDPQGIPLNLM-----PYVQV--AIGRR----- 220
QY 195 VNGAEVWVWGT-----GSPLEFLHVDLAD---ACVFLDRYSGLEHVNIGSGOEVT 245
Db 221 ---EALNVFGNDYDTEGTGVRDYIHVVDLAKGHIAALRKLKEQCGCRINYLTGTGYS 276
QY 246 IRELAEVLKVEVFGKLGWD-CTKPDGTPRKLMDSSKLA--SLGWTTPKVSRLDGLSQT 302
Db 277 VLQMVQAMEKVSQ--KKIPYKVVARREGDVAVCYANPSLAQBELGWTALGLDRMCDLW 334
QY 303 DWYLNK 308
Db 335 RWQKQN 340

RESULT 11
US-09-624-183-2
/ Sequence 2, Application US/09624183
/ Patent No. 6451577
/ GENERAL INFORMATION:
/ APPLICANT: JI et al.
/ TITLE OF INVENTION: HUMAN URIDINE DIPHOSPHATE GALACTOSE-4-EPIMERASE
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: HUMAN GENOME SCIENCES, INC.
/ STREET: 9410 KEY WEST AVE
/ CITY: ROCKVILLE
/ STATE: MD
/ COUNTRY: US
/ ZIP: 20850
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/624,183
/ FILING DATE: 24-Jul-2000
```

```
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/113,536
/ FILING DATE: JUL-10-98
/ APPLICATION NUMBER: 08/462,966
/ FILING DATE: JUN-05-95
/ APPLICATION NUMBER: PCT/US95/05785
/ FILING DATE: MAY-1195
/ ATTORNEY/AGENT INFORMATION:
/ NAME: JONATHAN L. KLEIN
/ REGISTRATION NUMBER: 41,119
/ REFERENCE/DOCKET NUMBER: PF162CID1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 301-610-5790
/ TELEFAX: 301-309-8439
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 348 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-624-183-2

Query Match 10.3%; Score 169.5; DB 4; Length 348;
Best Local Similarity 21.9%; Pred. No. 5.5e-10;
Matches 80; Conservative 58; Mismatches 135; Indels 93; Gaps 15;

QY 7 KIFVAGHRLGVSIAVRKLOEQGFTNLVLT-----HA 39
Db 4 KVLVTGAGYIGSHTVLELEAGYLPVVDNFHNAFRGGSLPESLRRVQELTGRSVEFE 63
QY 40 ELDLTQADVESFSEKFPVYVILAAKVGGIHA---NNTYPADFIGVNLQIOTNVHSA 96
Db 64 EMDILDQAGLQRLFKK---YSFMAVIHFAGLKAVGESVQKPLDYRVNLTGTIQLLEIM 119
QY 97 YEHGVKKLLPLGSSCIYPKPAQPIPEASALLTASLEP--TNEWATAKIAGIKTCQAYRI 154
Db 120 KAHGVKNLVFSSSATVYGNPQYLPLDEAHPGTGCTNPGYKSKFFIEMIRDL--COADKT 177
QY 155 QHGWDAI-----SGMPTNLYGNPNDNFPHNSHVLPAALMRFPHEAK 194
Db 178 ---WNAVLLRYFNPTGAHSGCIGEDPQGIPLNLM-----PYVQV--AIGRR----- 220
QY 195 VNGAEVWVWGT-----GSPLEFLHVDLAD---ACVFLDRYSGLEHVNIGSGOEVT 245
Db 221 ---EALNVFGNDYDTEGTGVRDYIHVVDLAKGHIAALRKLKEQCGCRINYLTGTGYS 276
QY 246 IRELAEVLKVEVFGKLGWD-CTKPDGTPRKLMDSSKLA--SLGWTTPKVSRLDGLSQT 302
Db 277 VLQMVQAMEKVSQ--KKIPYKVVARREGDVAVCYANPSLAQBELGWTALGLDRMCDLW 334
QY 303 DWYLNK 308
Db 335 RWQKQN 340

RESULT 12
PCT-US95-05785-2
/ Sequence 2, Application PCT/US9505785
/ GENERAL INFORMATION:
/ APPLICANT: JI, H, ET AL.
/ TITLE OF INVENTION: HUMAN UDP GALACTOSE-4-EPIMERASE
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
/ STREET: 6 BECKER FARM ROAD
/ CITY: ROSELAND
/ STATE: NEW JERSEY
/ COUNTRY: USA
/ ZIP: 07068
/ COMPUTER READABLE FORM:
```

RESULT 13  
US-09-392-772-6  
; Sequence 6, Application US/09392772  
; Patent No. 6346403  
; GENERAL INFORMATION:  
; APPLICANT: Famodu, Layo O.  
; APPLICANT: Antoni Rafaleki  
; TITLE OF INVENTION: Methionine Metabolic Enzymes  
; FILE REFERENCE: BB-1241  
; CURRENT APPLICATION NUMBER: US/09/392,772  
; CURRENT FILING DATE: 1999-09-07  
; EARLIER APPLICATION NUMBER: 60/099,519  
; EARLIER FILING DATE: 1998-09-08





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 04:00:21 ; Search time 54 Seconds  
(without alignments)  
769.892 Million cell updates/sec

Title: US-10-089-014-1

Perfect score: 1646

Sequence: 1 MSDKSAKIFVAGHGLVGS.....SLRDGLSQTYDWLKNVNCNR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: A\_Geneseq\_101002.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1646	100.0	312	22	AAG62614
2	1515	92.0	314	21	AAV54114
3	1353	82.2	299	21	AAG09919
4	1353	82.2	307	21	AAG09918
5	1353	82.2	359	21	AAG09917
6	1241	75.4	310	22	AAW00095
7	971	59.0	339	18	AAW21770
8	971	59.0	339	18	AAW21771
9	943.5	57.3	317	21	AAV81360
10	937.5	57.0	321	20	AAV29659

11	937.5	57.0	321	21	AAV54115	A GDP-4-keto-6-deo
12	915.5	55.6	323	20	AAW88319	E. coli O157 antiG
13	718.5	43.7	576	22	ABG18368	Novel human diagno
14	671	40.8	162	22	AAW00029	GDP-4-keto-6-deoxy
15	593.5	36.1	307	20	AAW88305	E. coli O111 antiG
16	550.5	33.4	597	22	ABG28981	Novel human diagno
17	550.5	33.4	693	22	ABG25460	Novel human diagno
18	500	30.4	244	22	ABG18364	Novel human diagno
19	468.5	28.5	321	23	ABG34139	Antibody productio
20	463.5	28.2	321	20	AAV28286	Amino acid sequenc
21	463.5	28.2	321	21	AAV54116	A GDP-4-keto-6-deo
22	451.5	27.4	314	21	AAV81361	Human GDP-fucose s
23	437.5	26.6	321	22	ABV59947	Drosophila melanog
24	430	26.1	260	22	ABG18365	Novel human diagno
25	425	25.8	1165	22	ABG00133	Novel human diagno
26	381	23.1	157	22	ABG28975	Novel human diagno
27	362.5	22.0	545	22	ABG11190	Novel human diagno
28	337.5	20.5	214	18	AAW20804	H. pylori cytoplasm
29	324	19.7	455	22	AAU28305	Novel human secret
30	313.5	19.0	325	21	AAW07566	Protein encoded by
31	271	16.5	380	21	AAW35197	zea mays protein f
32	271	16.5	420	21	AAW35196	zea mays protein f
33	263.5	16.0	377	21	AAW06774	Arabidopsis thalia
34	263.5	16.0	377	21	AAW14539	Arabidopsis thalia
35	263.5	16.0	423	21	AAW14538	Arabidopsis thalia
36	263.5	16.0	426	21	AAW06773	Arabidopsis thalia
37	262.5	15.9	336	21	AAW83788	S. fradiae tylosin
38	253	15.4	308	22	AAW6739	Putative P. abyss
39	249	15.1	311	22	AAW90108	C glutamicum prote
40	249	15.1	311	22	AAW79385	Corynebacterium gl
41	249	15.1	311	22	AAW79467	Corynebacterium gl
42	247	15.0	314	21	AAW35198	zea mays protein f
43	240	14.6	226	18	AAW33273	S. fradiae tylosin
44	239.5	14.6	313	21	AAW06775	Arabidopsis thalia
45	239.5	14.6	313	21	AAW14540	Arabidopsis thalia

# ALIGNMENTS

RESULT 1  
AAG62614  
ID AAG62614 standard; protein; 312 AA.  
XX AC  
AC AAG62614;  
XX  
DT 06-SEP-2001 (first entry)  
XX  
DE A thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase.  
XX  
KW GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase; GDP-L-fucose;  
KW sugar chain.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200138507-A1.  
XX  
PD 31-MAY-2001.  
XX  
PF 30-MAR-2000; 2000WO-JP02049.  
XX  
PR 19-NOV-1999; 99JP-0329045.  
XX  
(AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
XX  
PA Nakayama K, Jigami Y;  
XX  
XX WPI; 2001-381292/40.  
XX  
XX N-PSDB; AAH45734.  
XX  
XX Recombinant GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase  
XX derived from Arabidopsis for efficient production of GDP (guanine  
XX diphosphate)-L-fucose in vitro or in vitro

XX PS Claim 1; Page 24-26; 35pp; Japanese.

XX CC The present invention provides the protein and coding sequences of the

XX CC Arabidopsis thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-

CC reductase. The sequence can be used for the efficient production of

CC GDP-L-fucose, which is essential to the production of functionally

CC important sugar chains containing L-fucose. The present sequence is the

XX CC protein of the invention.

XX SQ Sequence 312 AA;

Query Match 100.0%; Score 1646; DB 22; Length 312;

Best Local Similarity 100.0%; Pred. No. 1.5e-159;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

DB 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

QY 61 VILAAKVGGIHANNTPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAPOP 120

DB 61 VILAAKVGGIHANNTPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAPOP 120

QY 121 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDIAISGMTNLYGPNDFHPENS 180

DB 121 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDIAISGMTNLYGPNDFHPENS 180

QY 181 HVLPMRRFHEAKVGAEEVVMGTGSPRLREFLHVDDLDADACVFLLDRIQ 240

DB 181 HVLPMRRFHEAKVGAEEVVMGTGSPRLREFLHVDDLDADACVFLLDRIQ 240

QY 241 GQVETIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDLSQ 300

DB 241 GQVETIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDLSQ 300

QY 301 TYDWYLNKVCNR 312

DB 301 TYDWYLNKVCNR 312

RESULT 2

AAY54114

ID AAY54114 standard; Protein; 314 AA.

AC AAY54114;

XX 27-MAR-2000 (first entry)

DT A GDP-4-keto-6-deoxy-D-mannose epimerase/reductase protein.

DE GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;

XX GDP-L-galactose; vitamin C; ascorbic acid; L-ascorbic acid;

KW ascorbic acid pathway enzyme; hexokinase; glucose phosphate isomerase;

KW phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;

KW GDP-D-mannose pyrophosphorylase; GDP-D-mannose:GDP-L-galactose epimerase;

KW GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;

KW L-galactono-gamma-lactone dehydrogenase; ester.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX WO9964618-A1.

PN 16-DEC-1999.

PD 26-MAY-1999; 99WO-US11576.

PF 08-JUN-1998; 98US-0088549.

PR 17-MAR-1999; 99US-0125073.

PR 18-MAR-1999; 99US-0125054.

XX (DCVB-) DCV INC DBA BIO-TECH RESOURCES.

XX

PI Berry A, Running JA, Severson DK, Burlingame RP;

XX WPI: 2000-105890/09.

DR N-PSDB; AA245315.

XX Production of ascorbic acid or esters, using microorganisms or plants

PT which have genetic modification in enzymes involved in the ascorbic

PT acid synthesis pathway -

XX Claim 27; Page 167-168; 187pp; English.

XX The present sequence represents a GDP-4-keto-6-deoxy-D-mannose epimerase/

CC reductase. The enzyme catalyses the conversion of GDP-D-mannose to

CC GDP-L-galactose. The enzyme can be modified, and used to produce

CC transgenic microorganisms, which can be used in fermentation techniques

CC to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is

CC modified to increase its action. Other ascorbic acid pathway enzymes

CC which may be used in the method of the invention include hexokinases,

CC glucose phosphate isomerases, phosphomannose isomerases,

CC phosphomannomutases, GDP-D-mannose pyrophosphorylases,

CC GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,

CC L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and

CC L-galactono-gamma-lactone dehydrogenases. The methods can be used for

CC the production of ascorbic acid or esters using microorganisms or plants.

XX SQ Sequence 314 AA;

Query Match 92.0%; Score 1515; DB 21; Length 314;

Best Local Similarity 92.5%; Pred. No. 3.7e-146;

Matches 298; Conservative 1; Mismatches 5; Indels 18; Gaps 5;

QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

DB 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

QY 61 VILAAKVGGIHANNTPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAPOP 120

DB 61 VILAAKVGGIHANNTPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAPOP 120

QY 121 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDIAISGMTNLYGPNDFHPENS 180

DB 121 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDIAISGMTNLYGPNDFHPENS 179

QY 181 HVLPMRRFHEAKVGAEEVVMGTGSPRLREFLHVDDLDADACVFLLDRIQ 231

DB 180 HVLPMRRFHEAKVGAEEVVMGTGSPRLREFLHVDDLDADACVFLLDRIQ 232

QY 232 -GLEHVNIGSGQVETIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTP 290

DB 233 RGLHVNIGSGQVETIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTP 292

QY 291 KVSRLDLSQTYDWYLNKVCNR 312

DB 293 KVSRLDLSQTYDWYLNKVCNR 314

RESULT 3

AAG09919

ID AAG09919 standard; Protein; 299 AA.

AC AAG09919;

XX 17-OCT-2000 (first entry)

DT Arabidopsis thaliana protein fragment SEQ ID NO: 8036.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1033405-A2.



XX	PD	06-SEP-2000.		PR	02-JUL-1999;	99US-0142055.
XX	PF	25-FEB-2000; 2000EP-0301439.		PR	06-JUL-1999;	99US-0142390.
XX	XX			PR	08-JUL-1999;	99US-0142803.
XX	XX			PR	09-JUL-1999;	99US-0142920.
XX	XX			PR	12-JUL-1999;	99US-0142977.
XX	XX			PR	13-JUL-1999;	99US-0143542.
XX	XX			PR	14-JUL-1999;	99US-0143624.
XX	XX			PR	15-JUL-1999;	99US-0144005.
XX	XX			PR	16-JUL-1999;	99US-0144085.
XX	XX			PR	16-JUL-1999;	99US-0144086.
XX	XX			PR	19-JUL-1999;	99US-0144325.
XX	XX			PR	19-JUL-1999;	99US-0144331.
XX	XX			PR	19-JUL-1999;	99US-0144332.
XX	XX			PR	19-JUL-1999;	99US-0144333.
XX	XX			PR	19-JUL-1999;	99US-0144334.
XX	XX			PR	19-JUL-1999;	99US-0144335.
XX	XX			PR	20-JUL-1999;	99US-0144352.
XX	XX			PR	20-JUL-1999;	99US-0144632.
XX	XX			PR	20-JUL-1999;	99US-0144632.
XX	XX			PR	21-JUL-1999;	99US-0144884.
XX	XX			PR	21-JUL-1999;	99US-0144884.
XX	XX			PR	21-JUL-1999;	99US-0145086.
XX	XX			PR	21-JUL-1999;	99US-0145088.
XX	XX			PR	22-JUL-1999;	99US-0145085.
XX	XX			PR	22-JUL-1999;	99US-0145087.
XX	XX			PR	22-JUL-1999;	99US-0145089.
XX	XX			PR	22-JUL-1999;	99US-0145192.
XX	XX			PR	23-JUL-1999;	99US-0145145.
XX	XX			PR	23-JUL-1999;	99US-0145218.
XX	XX			PR	23-JUL-1999;	99US-0145224.
XX	XX			PR	26-JUL-1999;	99US-0145276.
XX	XX			PR	27-JUL-1999;	99US-0145913.
XX	XX			PR	27-JUL-1999;	99US-0145918.
XX	XX			PR	27-JUL-1999;	99US-0145919.
XX	XX			PR	28-JUL-1999;	99US-0145951.
XX	XX			PR	02-AUG-1999;	99US-0146386.
XX	XX			PR	02-AUG-1999;	99US-0146388.
XX	XX			PR	02-AUG-1999;	99US-0146389.
XX	XX			PR	03-AUG-1999;	99US-0147038.
XX	XX			PR	04-AUG-1999;	99US-0147204.
XX	XX			PR	04-AUG-1999;	99US-0147302.
XX	XX			PR	05-AUG-1999;	99US-0147192.
XX	XX			PR	05-AUG-1999;	99US-0147260.
XX	XX			PR	06-AUG-1999;	99US-0147303.
XX	XX			PR	06-AUG-1999;	99US-0147416.
XX	XX			PR	09-AUG-1999;	99US-0147493.
XX	XX			PR	09-AUG-1999;	99US-0147935.
XX	XX			PR	10-AUG-1999;	99US-0148171.
XX	XX			PR	11-AUG-1999;	99US-0148319.
XX	XX			PR	12-AUG-1999;	99US-0148341.
XX	XX			PR	13-AUG-1999;	99US-0148565.
XX	XX			PR	16-AUG-1999;	99US-0149368.
XX	XX			PR	17-AUG-1999;	99US-0149175.
XX	XX			PR	18-AUG-1999;	99US-0149426.
XX	XX			PR	20-AUG-1999;	99US-0149722.
XX	XX			PR	20-AUG-1999;	99US-0149723.
XX	XX			PR	20-AUG-1999;	99US-0149929.
XX	XX			PR	23-AUG-1999;	99US-0149902.
XX	XX			PR	23-AUG-1999;	99US-0149930.
XX	XX			PR	25-AUG-1999;	99US-0150566.
XX	XX			PR	26-AUG-1999;	99US-0150884.
XX	XX			PR	27-AUG-1999;	99US-0151065.
XX	XX			PR	27-AUG-1999;	99US-0151066.
XX	XX			PR	27-AUG-1999;	99US-0151080.
XX	XX			PR	30-AUG-1999;	99US-0151303.
XX	XX			PR	31-AUG-1999;	99US-0151438.
XX	XX			PR	01-SEP-1999;	99US-0151930.
XX	XX			PR	07-SEP-1999;	99US-0152363.
XX	XX			PR	10-SEP-1999;	99US-0153070.
XX	XX			PR	13-SEP-1999;	99US-0153758.
XX	XX			PR	15-SEP-1999;	99US-0154018.
XX	XX			PR	16-SEP-1999;	99US-0154039.
XX	XX			PR	20-SEP-1999;	99US-0154779.

```
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.2%; Score 1353; DB 21; Length 299;
Best Local Similarity 89.5%; Pred. No. 1.2e-129;
Matches 255; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHRLVGSALVRKLOEQFTNLVLKTHAELDLTRQADVESFFSOEKPY 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 7 MLEKSAKIFVAGHRLVGSALVRKLOEQFTNLVLKTHAELDLTRQADVESFFATEKPY 66
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 VILAAKVGGIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKKLLFLGSSCIYKPAQP 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 67 VILAAKVGGIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKKLLFLGSSCIYKPAQP 126
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 IPESALLTASLEPTNEWYAIKTAGIKTCQAVRIQHWDAISGMPTNLGPNDFHPENS 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 127 IPESALLTGLEPTNEWYAIKTAGIKTCQAVRLQHWDAISGMPTNLGQDNDFHPENS 186
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 HVLPALMRPFHAKVNGASEVWVGSGPLREFLHVDDLDACVFLLDYRSGLEHVNIGS 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 HVLPALMRPFHAKVNGASEVWVGSGPLREFLHVDDLDACVFLMDQYSGFEHVNGS 246
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 241 GQEVITRELAELVKEVVGEGKLGWDCPTKPDGTPRKLMDSKSLAS 285
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 247 GVEVTIKELAELVKEVVGEGKLGWDDTKPDGTPRKLMDSKSLAS 291
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

## RESULT 4

AAG09918

ID AAG09918 standard; Protein; 307 AA.

XX AC

AAG09918;

XX DT

17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 8035.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 17-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 25-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 27-JUN-1999; 99US-0140891.  
PR 28-JUN-1999; 99US-0141287.  
PR 29-JUN-1999; 99US-0141842.  
PR 30-JUN-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 03-JUL-1999; 99US-0142390.  
PR 04-JUL-1999; 99US-0142803.  
PR 05-JUL-1999; 99US-0142820.  
PR 06-JUL-1999; 99US-0142877.  
PR 07-JUL-1999; 99US-0143542.  
PR 08-JUL-1999; 99US-0143624.  
PR 09-JUL-1999; 99US-0144085.  
PR 10-JUL-1999; 99US-0144086.  
PR 11-JUL-1999; 99US-0144325.  
PR 12-JUL-1999; 99US-0144331.  
PR 13-JUL-1999; 99US-0144332.  
PR 14-JUL-1999; 99US-0144333.  
PR 15-JUL-1999; 99US-0144334.  
PR 16-JUL-1999; 99US-0144335.  
PR 17-JUL-1999; 99US-0144352.  
PR 18-JUL-1999; 99US-0144632.  
PR 19-JUL-1999; 99US-0144884.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 23-JUL-1999; 99US-0145088.  
PR 24-JUL-1999; 99US-0145087.  
PR 25-JUL-1999; 99US-0145089.  
PR 26-JUL-1999; 99US-0145192.  
PR 27-JUL-1999; 99US-0145192.  
PR 28-JUL-1999; 99US-0145218.  
PR 29-JUL-1999; 99US-0145224.  
PR 30-JUL-1999; 99US-0145276.  
PR 01-AUG-1999; 99US-0145913.  
PR 02-AUG-1999; 99US-0145913.  
PR 03-AUG-1999; 99US-0145919.  
PR 04-AUG-1999; 99US-0145951.  
PR 05-AUG-1999; 99US-0145951.  
PR 06-AUG-1999; 99US-0146386.  
PR 07-AUG-1999; 99US-0146386.  
PR 08-AUG-1999; 99US-0146388.  
PR 09-AUG-1999; 99US-0147038.  
PR 10-AUG-1999; 99US-0147038.  
PR 11-AUG-1999; 99US-0147302.  
PR 12-AUG-1999; 99US-0147192.  
PR 13-AUG-1999; 99US-0147260.  
PR 14-AUG-1999; 99US-0147303.  
PR 15-AUG-1999; 99US-0147303.  
PR 16-AUG-1999; 99US-0147416.  
PR 17-AUG-1999; 99US-0147493.  
PR 18-AUG-1999; 99US-0147935.  
PR 19-AUG-1999; 99US-0148171.  
PR 20-AUG-1999; 99US-0148319.  
PR 21-AUG-1999; 99US-0148341.  
PR 22-AUG-1999; 99US-0148565.  
PR 23-AUG-1999; 99US-0148565.  
PR 24-AUG-1999; 99US-0148584.  
PR 25-AUG-1999; 99US-0149368.  
PR 26-AUG-1999; 99US-0149368.  
PR 27-AUG-1999; 99US-0149175.  
PR 28-AUG-1999; 99US-0149426.  
PR 29-AUG-1999; 99US-0149722.  
PR 30-AUG-1999; 99US-0149723.  
PR 01-SEP-1999; 99US-0149929.  
PR 02-SEP-1999; 99US-0149929.  
PR 03-SEP-1999; 99US-0149930.  
PR 04-SEP-1999; 99US-0150566.  
PR 05-SEP-1999; 99US-0150566.  
PR 06-SEP-1999; 99US-0150884.  
PR 07-SEP-1999; 99US-0151065.  
PR 08-SEP-1999; 99US-0151066.  
PR 09-SEP-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 02-SEP-1999; 99US-0152363.  
PR 03-SEP-1999; 99US-0153070.  
PR 04-SEP-1999; 99US-0153758.  
PR 05-SEP-1999; 99US-0154018.  
PR 06-SEP-1999; 99US-0154039.  
PR 07-SEP-1999; 99US-0154779.  
PR 08-SEP-1999; 99US-0155139.  
PR 09-SEP-1999; 99US-0155486.  
PR 10-SEP-1999; 99US-0155659.  
PR 11-SEP-1999; 99US-0156458.  
PR 12-SEP-1999; 99US-0156596.  
PR 13-SEP-1999; 99US-0157117.  
PR 14-SEP-1999; 99US-0157753.  
PR 15-SEP-1999; 99US-0157865.  
PR 16-SEP-1999; 99US-0158029.  
PR 17-SEP-1999; 99US-0158232.  
PR 18-SEP-1999; 99US-0158369.  
PR 19-SEP-1999; 99US-0159293.  
PR 20-SEP-1999; 99US-0159294.  
PR 21-SEP-1999; 99US-0159295.  
PR 22-SEP-1999; 99US-0159329.  
PR 23-SEP-1999; 99US-0159330.  
PR 24-SEP-1999; 99US-0159331.  
PR 25-SEP-1999; 99US-0159637.  
PR 26-SEP-1999; 99US-0159638.  
PR 27-SEP-1999; 99US-0159584.  
PR 28-SEP-1999; 99US-0160741.  
PR 29-SEP-1999; 99US-0160767.  
PR 30-SEP-1999; 99US-0160768.  
PR 01-OCT-1999; 99US-0160770.  
PR 02-OCT-1999; 99US-0160814.  
PR 03-OCT-1999; 99US-0160815.  
PR 04-OCT-1999; 99US-0160880.  
PR 05-OCT-1999; 99US-0160981.  
PR 06-OCT-1999; 99US-0160989.  
PR 07-OCT-1999; 99US-0161404.  
PR 08-OCT-1999; 99US-0161405.  
PR 09-OCT-1999; 99US-0161406.  
PR 10-OCT-1999; 99US-0161359.  
PR 11-OCT-1999; 99US-0161360.  
PR 12-OCT-1999; 99US-0161361.  
PR 13-OCT-1999; 99US-0161920.  
PR 14-OCT-1999; 99US-0161922.  
PR 15-OCT-1999; 99US-0161993.  
PR 16-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 1353; DB 21; Length 307;  
Best Local Similarity 89.5%; Pred. No. 1.3e-129;  
Matches 255; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSDSAKIFVAGHGLVGSIAIVRKLOEQFTNLVLTAEHDLTROADVESPQKPVY 60  
DB 15 MLESAKIFVAGHGLVGSIAIVRKLOEQFTNLVLTAEHDLTROADVESPQKPVY 74  
QY 61 VILAAKVGGIHANNTPADFIGVNLQIOTNVHSAYEHGVKLLFLGSSCIYPKFAPQ 120  
DB 75 VILAAKVGGIHANNTPADFIGVNLQIOTNVHSAYEHGVKLLFLGSSCIYPKFAPQ 134  
QY 121 IPESALLTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDIAISGMPTNLYGPNDFHPENS 180  
DB 135 IPESALLTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDIAISGMPTNLYGPNDFHPENS 194  
QY 181 HVLPAIMRRPFHEAKUNGABEVVWGSPLEFHLVDDIADACVFLLDYSLGHEVNVGS 240  
DB 195 HVLPAIMRRPFHEAKUNGABEVVWGSPLEFHLVDDIADACVFLLDYSLGHEVNVGS 254  
QY 241 GOEVTIRELAELVKEVVGEGKLGWDCTKPDGTPRKLMDSKIAS 285  
DB 255 GVEVTIRELAELVKEVVGEGKLGWDCTKPDGTPRKLMDSKIAS 299

RESULT 5  
AAG09917  
ID AAG09917 standard; Protein; 359 AA.  
XX AC AAG09917;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 8034.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 1353; DB 21; Length 359;  
 Best Local Similarity 89.5%; Pred. No. 1.6e-129;  
 Matches 255; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHGLVSAIVRKLQEQFTNLVLKTHAELDLTROADVESFFSOEKPYV 60  
 Db 67 MLEKSAKIFVAGHGLVSAIVRKLQEQFTNLVLKTHAELDLTROADVESFFSOEKPYV 126

QY 61 VILAAKVGGIHANNYPADFTGNLQIQTNVIHSAHYGVKKLLPLGSSCIYKFPAPOP 120.  
 Db 127 VILAAKVGGIHANNYPADFTGNLQIQTNVIHSAHYGVKKLLPLGSSCIYKFPAPOP 186

QY 121 IPESALLTASPTNEWYAIKTAGIKTCQAVRIQHWDAISGMPTNLGPNNDHPENS 180  
 Db 187 IPESALLTGPLEFTNEWYAIKTAGIKTCQAVRIQHWDAISGMPTNLGPNNDHPENS 246

QY 181 HVLPMRRFHEAKVNGAEVVVWGSGPLRFLHVDDLDACVFLDDRYSGLEHVNIGS 240  
 Db 247 HVLPMRRFHEAKVNGAEVVVWGSGPLRFLHVDDLDACVFLDDRYSGLEHVNIGS 306  
 QY 241 GOEVTIRELAELAVKEVGFEGKLGWDTKPDCTPRKLMDSKSLAS 285  
 Db 307 GVEVTIKELAEVLKEVGFEGKLGWDTTTPRKLMDNSKSLAS 351

RESULT 6

AAM00095  
 ID AAM00095 standard; Protein; 310 AA.

XX AAM00095;

XX 28-SEP-2001 (first entry)

DE GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase-4-reductase sequence #163.

KW Moss; carbohydrate metabolism related protein; CMRP; sugar; cofactor;  
 fine chemical production; carbohydrate; polysaccharide.

XX Physcomitrella patens.

XX WO200144476-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-EP12697.

XX 16-DEC-1999; 99US-0171101.

XX (BADI ) BASF PLANT SCI GMBH.

PI Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff P;  
 PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;

XX WPI; 2001-398155/42.

XX N-PSDB; AAH88781.

PT Novel moss nucleic acid molecules encoding a carbohydrate metabolism  
 related protein useful for modulating production of fine chemicals such  
 as carbohydrates, cofactors and enzymes from microorganisms and plants

XX Claim 30; Page 127; 133pp; English.

CC This invention relates to nucleic acid molecules AAH88708 - AAH88796  
 isolated from Physcomitrella patens (a moss), which encode carbohydrate  
 metabolism related proteins (CMRP) represented in AAM00022 - AAM00110.  
 CC Included in the invention is a vector containing the CMRP cDNA, and a  
 host cell transformed with the vector. The host cell (a microorganism,  
 CC Corynebacterium or Brevibacterium, moss or algae or a plant cell) is  
 useful for producing a fine chemical such as carbohydrates, cofactors  
 and/or enzymes. The nucleic acid molecules are suitable for modifying a  
 carbohydrate production system in a host, e.g., microorganisms and  
 CC plants. They are also useful to identify those DNA sequences and enzymes  
 in other species which are useful to modify the biosynthesis of starch,  
 CC cell wall polysaccharides and sugars. The nucleic acid molecules may be  
 CC utilised in the genetic engineering of Corynebacterium glutamicum and the  
 CC related Brevibacterium species and Acetobacter xylinum and Chlorella to  
 make it a better or more efficient producer of one or more fine  
 CC chemicals. Mutagenesis of one or more CMRPs may also result in CMRPs  
 CC having altered activities which indirectly impact the production of one  
 CC or more desired fine chemicals from plants. Primers AAH88705 - AAH8707  
 CC are used in the sequencing of the CMRP cDNA sequences of the invention.

SQ Sequence 310 AA;

Query Match 75.4%; Score 1241; DB 22; Length 310;

Best Local Similarity 75.5%; Pred. No. 3.7e-118;

Matches 231; Conservative 37; Mismatches 36; Indels 2; Gaps 1;





XX AC AAY29659;  
XX 03-NOV-1999 (first entry)  
XX Escherichia coli YEF B amino acid sequence.  
DE Escherichia coli; YEF B; expression vector; enzymatic conversion;  
KW guanosine diphosphate-4-keto-6-deoxymannose; GDP-mannose; GDP-fucose;  
KW fucosylated oligosaccharide; enzymatic fucosyl transfer; glycoprotein;  
KW glycolipid; insulin; growth hormone; tissue plasminogen activator;  
KW interleukin; viral antigen; therapeutic; diagnosis; food.  
XX Escherichia coli.  
XX WO9936555-A1.  
XX 22-JUL-1999.  
XX 15-JAN-1999; 99WO-US00893.  
XX 14-JAN-1999; 99US-0071076.  
XX 13-JAN-1998; 98US-0071076.  
XX (CYTE-) CYTEL CORP.  
XX Sjoberg ER;  
XX WPI; 1999-527202/44.  
XX New vector expressing an enzyme that converts guanosine  
PT diphosphate-4-keto-6-deoxymannose to GDP-fucose, used to prepare  
PT fucosylated oligosaccharides  
XX Claim 41; Page 77; 79pp; English.  
XX The present invention describes an expression vector comprising a  
CC promoter linked to a nucleic acid (I) that encodes a prokaryotic enzyme  
CC (II) having both epimerase and reductase activity, for the catalysis of  
CC conversion of GDP (guanosine diphosphate)-4-keto-6-deoxymannose (III) to  
CC GDP-fucose (IV), where the vector lacks an Escherichia coli wcaI coding  
CC region. An expression vector is specifically claimed where (I) encodes  
CC Escherichia coli YEF B. The present sequence represents the E. coli  
CC YEF B protein. (II) is used for production of (IV) which is then used to  
CC prepare fucosylated oligosaccharides (A) by enzymatic fucosyl transfer,  
CC e.g. to modify oligosaccharide components of glycoproteins or  
CC glycolipids, such as insulin, human or bovine growth hormones, tissue  
CC plasminogen activator, interleukins and viral antigens, or chimeric  
CC products such as immunoglobulin derivatives. (A) are variously useful as  
CC therapeutic and diagnostic agents and in foods. Combining two activities  
CC in a single enzyme simplifies the process, allowing efficient synthesis  
CC of complex fucosylated oligosaccharides in a single reaction vessel from  
CC readily available starting materials. The method is suitable for large  
CC scale synthesis, e.g. 0.2 kg batches. (II) can be expressed efficiently  
CC in prokaryotic cells (contrast similar mammalian enzymes).  
XX Sequence 321 AA;  
Query Match 57.0%; Score 937.5; DB 20; Length 321;  
Best Local Similarity 56.9%; Pred. No. 4.1e-87;  
Matches 177; Conservative 58; Mismatches 67; Indels 9; Gaps 1;  
7 KIFVAGHRLVGSIAVRKLEQCFNLVLKTHAELDLTRQADVESPFSEKPVYVTLAA 66  
5 RVFIAGHGMVGSIAIRQLQEQDVELVLRDELNLLDSRAVDFPAGERIDQVYVTLAA 64  
67 KVGGIHANNYPADFIQVNLQIQTNVHSAHEYKVKLLFLGSSCIYKPFAPQIPESAL 126  
65 KVGGIHANNYPADFIQVNNMIESNIIHAHQNDVKNKLLFLGSSCIYKPLAKQPAESEL 124  
127 LTASLEPTNEWYAIKIGTKTCOAYRIQHGDAISGMFTNLYGPDNPHNSHVPAL 186  
125 LGTLEPTNEPYAIAKIGKLCESYRQYGRDYRSWMTNLYGPDNPHNSHVPAL 184

QY 187 MRRFHEAKVNGAEVVMVGTSPLREFLHVDDLADAC-----VFLLDRYSGLHVN 237  
DB 185 LRRFHEATAQNAPODVVMVGSGTPEFLHVDDMAAASIHVMELAHEVWLENTPQMLSHN 244  
QY 238 IGSQGEVITRELAEVLKEVVGEGKLGWDCTKPGCTPRKLMDSKSLASLGWTPKVSRLRG 297  
DB 245 VGTGVDCTIRDVAQITAKVVGKGRWFDASKPDCTPRKLLDVTLRHLQLGWYHEISLEAG 304  
QY 298 LSQTYDWMVKLN 308  
DB 305 LASTYQWFLEN 315  
RESULT 11  
AAY54115  
ID AAY54115 standard; Protein; 321 AA.  
XX AAY54115;  
XX 27-MAR-2000 (first entry)  
DE A GDP-4-keto-6-deoxy-D-mannose epimerase/reductase protein.  
XX GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;  
KW GDP-L-galactose; Vitamin C; ascorbic acid; L-ascorbic acid;  
KW ascorbic acid pathway enzyme; hexokinase; Glucose phosphate isomerase;  
KW phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;  
KW GDP-D-mannose pyrophosphorylase; GDP-D-mannose:GDP-L-galactose epimerase;  
KW GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;  
KW L-galactono-gamma-lactone dehydrogenase; ester.  
XX Escherichia coli.  
XX WO9964618-A1.  
XX 16-DEC-1999.  
XX 26-MAY-1999; 99WO-US11576.  
XX 08-JUN-1998; 98US-0088549.  
XX 17-MAR-1999; 99US-0125073.  
XX 18-MAR-1999; 99US-0125054.  
XX (DCVB-) DCV INC DBA BIO-TECH RESOURCES.  
XX Berry A, Running JA, Severson DK, Burlingame RP;  
WPI; 2000-105890/09.  
N-PSDB; AA245316.  
Production of ascorbic acid or esters, using microorganisms or plants  
PT which have genetic modification in enzymes involved in the ascorbic  
PT acid synthesis pathway -  
XX Claim 27; Page 170-171; 187pp; English.  
XX The present sequence represents a GDP-4-keto-6-deoxy-D-mannose epimerase/  
CC reductase. The enzyme catalyses the conversion of GDP-D-mannose to  
CC GDP-L-galactose. The enzyme can be modified, and used to produce  
CC transgenic microorganisms, which can be used in fermentation techniques  
CC to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is  
CC modified to increase its action. Other ascorbic acid pathway enzymes  
CC which may be used in the method of the invention include hexokinases,  
CC glucose phosphate isomerases, phosphomannose isomerases,  
CC phosphomannomutases, GDP-D-mannose pyrophosphorylases,  
CC GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,  
CC L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and  
CC L-galactono-gamma-lactone dehydrogenases. The methods can be used for  
CC the production of ascorbic acid or esters using microorganisms or plants.  
XX Sequence 321 AA;  
SQ



	Qy	187	MRRFHEAKVNGAEVVWGTGTPUREFLHVDDLDADACVFLLDR-YS-----GLEHVN	237
		:	:    :    :	:
		:	:    :    :	:
		:	:    :    :	:
	Dd	187	LRFPHDVAENNSPNVNWGSSTGRKREFLHVDDMASASIYVMEMPDYIQWKNTKMVLSHIN	246

xx This is the amino acid sequence of a protein encoded by an open  
 CC reading frame (ORF8) of a gene cluster (see AAX06749) involved in  
 CC the biosynthesis of the *Escherichia coli* O157 O antigen. The  
 CC protein shows homology to the *E. coli* wcaG colanic acid capsule  
 CC protein. The use of nucleic acids derived from assembly and  
 CC transport genes, particularly wbd (transferase), wxx (flippase)  
 CC and wzy (polymerase) genes, within O antigen gene clusters

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 20; SEQ ID NO 48727; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 576 AA;

Query Match 43.7%; Score 718.5; DB 22; Length 576;

Best Local Similarity 58.2%; Pred. No. 2.3e-64;

Matches 142; Conservative 38; Mismatches 55; Indels 9; Gaps 1;

QY 23 RKLQEGFTNLVLTAEHLDTQADVESFQKPVYVILAAKVGGIHANNYPADFI 82

Db 293 RQLQEGDVELVLTDELNLLDSRAVHDFASERIDQVYLAARKVGIVANNYPADFI 352

QY 83 GVNLOQTQVIHSAHYGVKLLFLGSSCTYPKFAQPIPESSALLTASLEPTNEWAIK 142

Db 353 YQNMIESIIHAHQNDVKNLLFLGSSCTYPKLAQPMASSELQGTLEPTNEPYAIK 412

QY 143 IAGIKTCQAVRIQHGWDATSGMPTNLYGPNDFHPSNHLVLPALRRFHEAKVNGAEVY 202

Db 413 IAGIKLCEINRQIGRDYRSVMPTNLYGPHDNFPHSNHVPALLRRFHEATAQNAPDV 472

QY 203 VMTGSPFLREFLHVDDADAC-----VFLLDRYSGLEHVNIGSGQEVTIRELAEVL 253

Db 473 VMSGTPMREFLHVDDMAAAS IHVMELAEHVEWLENTPMLSHINVTGVDCTIRELAQTI 532

QY 254 KEVV 257

Db 533 GKVV 536

RESULT 14

AAW00029

ID AAW00029 standard; Protein; 162 AA.

XX AAW00029;

XX 28-SEP-2001 (first entry)

XX GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase-4-reductase cDNA sequence #97.

XX Moss; carbohydrate metabolism related protein; CWRP; sugar; cofactor;

KW fine chemical production; carbohydrate; polysaccharide.

XX Physcomitrella patens.

OS WO200144476-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000NO-EP12697.

XX 16-DEC-1999; 99US-0171101.

XX (BADI ) BASF PLANT SCI GMBH.

PA Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;

PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;

XX WPI; 2001-398155/42.

DR N-PSDB; AAH88715.

XX Novel moss nucleic acid molecules encoding a carbohydrate metabolism

PT related protein useful for modulating production of fine chemicals such

as carbohydrates, cofactors and enzymes from microorganisms and plants

XX Claim 30; Page 114; 133pp; English.

XX This invention relates to nucleic acid molecules AAH88708 - AAH88796

isolated from Physcomitrella patens (a moss), which encode carbohydrate

metabolism related proteins (CWRP) represented in AAM00022 - AAM00110.

CC Included in the invention is a vector containing the CWRP cDNA, and a

CC host cell transformed with the vector. The host cell (a microorganism,

CC Corynebacterium or Brevibacterium, moss or algae or a plant cell) is

CC useful for producing a fine chemical such as carbohydrates, cofactors

CC and/or enzymes. The nucleic acid molecules are suitable for modifying a

CC carbohydrate production system in a host, e.g., microorganisms and

CC plants. They are also useful to identify those DNA sequences and enzymes

CC in other species which are useful to modify the biosynthesis of starch

CC utilised in the genetic engineering of Corynebacterium glutamicum and the

CC related Brevibacterium species and Acetobacter xylinum and Chlorella to

CC make it a better or more efficient producer of one or more fine

CC chemicals. Mutagenesis of one or more CWRPs may also result in CWRPs

CC having altered activities which indirectly impact the production of one

CC or more desired fine chemicals from plants. Primers AAH88705 - AAH88707

CC are used in the sequencing of the CWRP cDNA sequences of the invention.

XX Sequence 162 AA;

Query Match 40.8%; Score 671; DB 22; Length 162;

Best Local Similarity 74.8%; Pred. No. 2.7e-60;

Matches 119; Conservative 25; Mismatches 15; Indels 0; Gaps 0;

QY 150 QAYRIQHGWDAISGMTNLYGPNDFHPSNHLVLPALRRFHEAKVNGAEVYVMTGSP 209

Db 1 QAYRIQYNFDALSGMPTNLYGPHDNFPHPSNHLVLPALRRFHEAKVNGAEVYVMTGSP 60

QY 210 LREFLHVDDADACVFLDRYSGLEHVNIGSGQEVTIRELAEVLKVEVGFEGKLGWDTCK 269

Db 61 FREFLHVDDLAETVFLQNTYSAHEHVNMGSGSEVSIKELAEVYVGFQGLTWDTSK 120

QY 270 PDGTPKRLMDSKSLASLGWTPKVSLRDCLSQTYDHYLKN 308

Db 121 PDGTPKRLIDSSKLANMGWQARIPLKEGLAETRYKWCYN 159

RESULT 15

AAW88305

ID AAW88305 standard; Protein; 307 AA.

XX AAW88305;

XX 26-APR-1999 (first entry)

XX E. coli O111 antigen gene cluster ORF6 (wbdJ) protein product.

DE O antigen; O111 antigen; wbdJ gene; diarrhoea;

KW haemorrhagic colitis; diagnosis.

XX Escherichia coli.

XX WO980531-A1.

XX 12-NOV-1998.





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2003, 02:39:05 ; Search time 1673 Seconds  
(without alignments)

9060.954 Million cell updates/sec

Title: US-10-089-014-2

Perfect score: 936

Sequence: 1 atgtctgacaaatctgccaa.....tgaagaatgtttgcaaccca 936

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527	56.3	627	9 AU237351	AU237351 AU237351
C 2	513	54.8	537	17 CNS000P9U	AL084240 Arabidops
C 3	436.2	46.6	441	17 B22510	B22510 F20D2TF IGF
4	417.8	44.6	856	12 BF276275	BF276275 CA_Eb002
5	416.6	44.5	852	12 BF278112	BF278112 CA_Eb003
C 6	408.2	43.6	1439	11 AY105324	AY105324 Zea mays

7	386	41.2	618	17	A0841772
8	354.8	37.9	652	10	AW041037
9	354.4	37.9	731	17	BH173899
10	352.2	37.6	566	17	BH173840
11	350.6	37.5	752	14	BQ116267
12	339.8	36.3	687	12	BG440809
13	338.6	36.2	614	9	AI897511
14	326	34.8	664	9	AI726948
15	322.2	34.4	590	9	AI487671
16	314.8	33.6	539	13	BI424065
17	308.6	33.0	547	10	BE458430
18	308.6	33.0	665	14	BQ405379
19	299.8	32.0	649	10	AW761509
20	295.2	31.5	739	10	BE055323
21	295	31.5	513	13	BI498273
22	290.8	31.1	497	10	AW507622
23	290.2	31.0	512	10	BE459038
24	287.8	30.7	748	17	BH547992
C 25	282	30.1	667	14	BQ411981
26	277.8	29.7	547	10	AW906983
27	268.8	28.7	302	14	Z35041
28	265.4	28.4	539	13	BM110048
29	258.6	27.6	529	10	BE343057
30	256.6	27.4	737	9	AF109667
C 31	256.4	27.4	890	17	BI9210
32	254	27.1	522	9	AI726475
33	253.6	27.1	757	10	BE660172
34	250.4	26.8	481	13	BI433478
35	246.8	26.4	603	13	BI893509
36	241.2	25.8	669	14	BM779643
37	240.6	25.7	472	10	BE463183
38	240	25.6	396	12	BG405837
C 39	239.6	25.6	648	12	BE823790
C 40	235.8	25.2	239	17	BH851675
41	230.2	24.6	541	13	BI699441
42	230	24.6	622	10	BE440335
43	229.6	24.5	570	13	BM371263
44	228.6	24.4	657	14	BQ408320
C 45	227	24.3	437	10	AV560875

## ALIGNMENTS

### RESULT 1

AU237351 627 bp mRNA linear EST 01-APR-2002  
LOCUS AU237351 RAF116 Arabidopsis thaliana cDNA clone RAF116-16-116 5',  
DEFINITION AU237351 mRNA sequence.  
ACCESSION AU237351  
VERSION AU237351.1 GI:19876520  
KEYWORDS EST.  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 627)  
AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.  
TITLE Large scale analysis of Arabidopsis full-length cDNA  
JOURNAL Unpublished (2002)  
COMMENT Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: mseki@tc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

A0841772 TL34022 S  
AW041037 EST283901  
BH173899 B216/R RF  
BH173840 A890/R RF  
BQ116267 EST601830  
BG440809 GA\_Ea001  
AI897511 EST26954  
AI726948 BNLGH1690  
AI487671 EST45993  
BI424065 sah5d08.  
BE458430 EST413722  
BQ405379 GA\_Ed008  
AW761509 sl58e03.y  
BE055323 GA\_Ea000  
BI498273 sag18b01.  
AW507622 sl42b09.y  
BE459038 EST414330  
BH547992 BORMI23TR  
BQ411981 GA\_Ed004  
AW906983 EST343210  
Z35041 ATTS3680 G1  
BM110048 EST57584  
BE343057 EST395901  
AF109667 AF109667  
BI9210 F8H3-17 IGF  
AI726475 BNLGH1595  
BE660172 831 GmaxS  
BI433478 EST536239  
BI893509 sai67b09.  
BM779643 EST590219  
BE463183 EST354406  
BG405837 sac38e01.  
BE823790 GM700022A  
BH851675 SALK\_0733  
BI699441 sag37a06.  
BE440335 sp44c03.y  
BM371263 EBma08\_SQ  
BQ408320 GA\_Ed000  
AV560875 AV560875

and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pluescript vector. Please visit our web site (<http://www.genoscope.cns.fr>) for further details.

```
FEATURES
  source      Location/Qualifiers
    1..627
      /organism="Arabidopsis thaliana"
      /db_xref="taxon:3702"
      /clone="RAFL16-16-116"
      /lab_host="RAFL16"
      /lab_host="DH10B"
      /notes="Site 1: BamHI; Site 2: SalI; dark-grown"
      /note="end : 189"
BASE COUNT      151 a 159 c 125 g 189 t 3 others
ORIGIN
Query Match      56.3%; Score 527; DB 9; Length 627;
Best Local Similarity 99.4%; Pred. No. 3.2e-144;
Matches 527; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 60
Db 98 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 157
QY 61 ATTGTCGCAAGCTTCAGGAACAAGGTTTCCCAATCTCGTCTTAAACACACACGCCGAG 120
Db 158 ATTGTCGCAAGCTTCAGGAACAAGGTTTCCCAATCTCGTCTTAAACACACACGCCGAG 217
QY 121 CTTGATCTACTCGTCAAGCGGATGTGAATCTTCTTTTCAAGAAAGCCAGTTTAT 180
Db 218 CTTGATCTACTCGTCAAGCGGATGTGAATCTTCTTTTCAAGAAAGCCAGTTTAT 277
QY 181 GTAATCTAGCAGCAGTAAAGTTGGTGTATTCACGCTTAACAACACCTATCTCGTGAT 240
Db 278 GTAATCTAGCAGCAGTAAAGTTGGTGTATTCACGCTTAACAACACCTATCTCGTGAT 337
QY 241 TTCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTTGCATATGAGCACGGT 300
Db 338 TTCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTTGCATATGAGCACGGT 397
QY 301 GTGAAGAAGTTCTCTTCTTGATCATCTGATCATCTGATTAACCTTAAATTTGCTCTAGCCA 360
Db 398 GTGAAGAAGTTCTCTTCTTGATCATCTGATCATCTGATTAACCTTAAATTTGCTCTAGCCA 457
QY 361 ATTCCTGAGTCTGTTTGAACAGCATCTTGAACCAACTTAATGAGTGTATGCTATT 420
Db 458 ATTCCTGAGTCTGTTTGAACAGCATCTTGAACCAACTTAATGAGTGTATGCTATT 517
QY 421 GCTAAGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGGATGGGATGCA 480
Db 518 GCTAAGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGGATGGGATGCA 577
QY 481 ATCTCTGGATCGCTACTAATCTCTATGCTCTATGCTCTATGACAAATTTCCACCC 530
Db 578 ATCTCTGGATCGCTACTAATCTCTATGCTCTATGCTCTATGACAAATTTCCANCC 627
```

```
RESULT 2
CNS00P9U/c      537 bp DNA linear GSS 28-JUN-1999
LOCUS      Arabidopsis thaliana genome survey sequence T7 end of BAC F8H3 of
DEFINITION      ICF library from strain Columbia of Arabidopsis thaliana, genomic
ACCESSION      AL084240
VERSION      AL084240.1 GI:5285380
KEYWORDS      GSS.
SOURCE      Arabidopsis thaliana.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 537)
REFERENCE      Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P.,
AUTHORS
```

```
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 537)
AUTHORS      Genoscope.
TITLE      Direct Submission
JOURNAL      Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
              - Web : www.genoscope.cns.fr)
FEATURES
  source      Location/Qualifiers
    1..537
      /organism="Arabidopsis thaliana"
      /strains="Columbia"
      /db_xref="taxon:3702"
      /clone="F8H3"
      /clone.lib="IGF"
      /note="end : 17"
BASE COUNT      164 a 110 c 131 g 132 t
ORIGIN
Query Match      54.8%; Score 513; DB 17; Length 537;
Best Local Similarity 99.8%; Pred. No. 3.9e-140;
Matches 524; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 60
Db 524 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 465
QY 61 ATTGTCGCAAGCTTCAGGAACAAGGTTTCCCAATCTCGTCTTAAACACACACGCCGAG 120
Db 464 ATTGTCGCAAGCTTCAGGAACAAGGTTTCCCAATCTCGTCTTAAACACACACGCCGAG 405
QY 121 CTTGATCTACTCGTCAAGCGGATGTGAATCTTCTTTTCAAGAAAGCCAGTTTAT 180
Db 404 CTTGATCTACTCGTCAAGCGGATGTGAATCTTCTTTTCAAGAAAGCCAGTTTAT 345
QY 181 GTAATCTAGCAGCAGTAAAGTTGGTGTATTCACGCTTAACAACACCTATCTCGTGAT 240
Db 344 GTAATCTAGCAGCAGTAAAGTTGGTGTATTCACGCTTAACAACACCTATCTCGTGAT 285
QY 241 TTCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTTGCATATGAGCACGGT 300
Db 284 TTCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTTGCATATGAGCACGGT 225
QY 301 GTGAAGAAGTTCTCTTCTTGATCATCTGATCATCTGATTAACCTTAAATTTGCTCTAGCCA 360
Db 224 GTGAAGAAGTTCTCTTCTTGATCATCTGATCATCTGATTAACCTTAAATTTGCTCTAGCCA 165
QY 361 ATTCCTGAGTCTGTTTGAACAGCATCTTGAACCAACTTAATGAGTGTATGCTATT 420
Db 164 ATTCCTGAGTCTGTTTGAACAGCATCTTGAACCAACTTAATGAGTGTATGCTATT 105
QY 421 GCTAAGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGGATGGGATGCA 480
Db 104 GCTAAGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGGATGGGATGCA 45
QY 481 ATCTCTGGATCGCTACTAATCTCTATGCTCTATGCTCTATGACAAATTTTC 525
Db 44 ATCTCTGGATCGCTACTAATCTCTATGCTCTATGACAAATTTTC 1
```

```
RESULT 3
B22510/c      441 bp DNA linear GSS 10-OCT-1997
LOCUS      B22510 ICF Arabidopsis thaliana genomic clone F20D2, DNA sequence.
DEFINITION      B22510
ACCESSION      B22510
VERSION      B22510.1 GI:2508240
KEYWORDS      GSS.
SOURCE      thale cress.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 441)
REFERENCE
```

**AUTHORS** Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Coffey, E., Golden, K., Johnson, K., Adams, M.D. and Venter, J.C.  
**TITLE** A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing  
**JOURNAL** Unpublished (1997)  
**COMMENT** Other\_GSSs: F20D2TR  
 Contact: Steve Rounsley  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: rounsley@tigr.org  
 Seq primer: M13-21  
 Class: BAC ends  
 High quality sequence stop: 441.  
**FEATURES**  
 source  
 1..441  
 /organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="F20D2"  
 /clone\_lib="IGF"  
 /sex="Hermaphrodite"  
 /note="Vector: Bel0BACII; Site\_1: EcoRI; Site\_2: EcoRI;  
 Produced by Thomas Altmann"  
**BASE COUNT** 133 a 87 c 108 g 113 t  
**ORIGIN**

Query Match 46.6%; Score 436.2; DB 17; Length 441;  
 Best Local Similarity 99.3%; Pred. No. 1.6e-117;  
 Matches 438; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 94 AATCTCGTTCTTAAACACACGCCGAGCTTGATCTCAGTCGCAAGCCGATGTGAATCC 153  
 DB |||||  
 QY 154 TTCTTTTCTCAAGAGAGCCAGTTTATGTAATCCTAGCAGCAGCTAAAGTTGGTGTATT 213  
 DB |||||  
 QY 381 TTCTTTTCTCAAGAGAGCCAGTTTATGTAATCCTAGCAGCAGCTAAAGTTGGTGTATT 322  
 DB |||||  
 QY 214 CAGGCTAACACACCTATCTCTGATTTCAATGGTGTCATCTCCAGATTGAGCAAT 273  
 DB |||||  
 QY 321 CAGGCTAACACACCTATCTCTGATTTCAATGGTGTCATCTCCAGATTGAGCAAT 262  
 DB |||||  
 QY 274 GTGATCCACTCTGCATATGAGCAGCGTGGAAGAGCTTCTCTCTGGATCATCTGC 333  
 DB |||||  
 QY 261 GTGATCCACTCTGCATATGAGCAGCGTGGAAGAGCTTCTCTCTGGATCATCTGC 202  
 DB |||||  
 QY 334 ATTTACCCCTAAATTTGCTCCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCATCGCTT 393  
 DB |||||  
 QY 201 ATTTACCCCTAAATTTGCTCCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCATCGCTT 142  
 DB |||||  
 QY 394 GAACCAACTAATGAGTGGTATGCTATTGTAAGATCGCTGGGATTAAGCTTGTAGGCT 453  
 DB |||||  
 QY 141 GAACCAACTAATGAGTGGTATGCTATTGTAAGATCGCTGGGATTAAGCTTGTAGGCT 82  
 DB |||||  
 QY 454 TATAGGATTCAGCAGCGATGGGATGCAATCTCTGGCATCGCTTAACTCTATGTCCT 513  
 DB |||||  
 QY 81 TATAGGATTCAGCAGCGATGGGATGCAATCTCTGGCATCGCTTAACTCTATGTCCT 22  
 DB |||||  
 QY 514 AATGACAAATTTCCACCCGGAG 534  
 DB |||||  
 QY 21 AATGACAAATTTCCACCCGGAG 1

**RESULT 4**  
**BF276275**  
**LOCUS** BF276275 856 bp mRNA linear EST 07-MAR-2001  
**DEFINITION** GA\_Eb0028C16f Gossypium arboreum 7-10 dpa fiber library Gossypium  
**ACCESSION** arboreum cDNA clone GA\_Eb0028C16f, mRNA sequence.  
**VERSION** BF276275.1 GI:11207345

**KEYWORDS** EST.  
**SOURCE** Gossypium arboreum.  
**ORGANISM** Gossypium arboreum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
**REFERENCE** 1 (bases 1 to 856)  
**AUTHORS** Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry  
 D., Wood, T.C., Leslie, A. and Wilkins, T.A.  
**TITLE** An integrated analysis of the genetics, development, and evolution  
 of the cotton fiber  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: TAATACGACTCATATAGG  
 High quality sequence stop: 712.  
**FEATURES**  
 source  
 1..856  
 /organism="Gossypium arboreum"  
 /strain="ARA"  
 /cultivar="8400"  
 /db\_xref="taxon:29729"  
 /clone="GA\_Eb0028C16f"  
 /clone\_lib="Gossypium arboreum 7-10 dpa fiber library"  
 /tissue\_type="Fibers isolated from bolls harvested 7-10  
 dpa"  
 /lab\_host="E. coli"  
 /note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
**BASE COUNT** 213 a 217 c 208 g 217 t 1 others  
**ORIGIN**

Query Match 44.6%; Score 417.8; DB 12; Length 856;  
 Best Local Similarity 70.2%; Pred. No. 6.4e-112;  
 Matches 560; Conservative 0; Mismatches 238; Indels 0; Gaps 0;  
 QY 2 TCTCTGACAAATCTCCAAATCTTCGTCGGGTCATCGTGTGGTTCGATCGCCA 61  
 DB |||||  
 QY 51 TCTCGAAACAATCAGCGAAGATCTTCGTAGCAGGCGATCGAGGCTTAGTGTTCGCCA 110  
 DB |||||  
 QY 62 TTGTCGCGAAGCTTCAGGAACAAGGTTTCAACCAATCTCTTCTTAAACAACACGCGAGC 121  
 DB |||||  
 QY 111 TAGTTCGTAACTCAGTCCCTAGCTTTACCAACCTCTACTCCGACCCATCGCATC 170  
 DB |||||  
 QY 122 TTGATCTCACTCGTCAGCCGATGTAATCCTTTCTTCAAGAGAGCCAGTTTATG 181  
 DB |||||  
 QY 171 TGGACCTTACTCGCCAATCCGACGTCGAATCCTTCTTCGCCGATGAGAAACCTCACTATG 230  
 DB |||||  
 QY 182 TAATCCTAGCAGCAGTAAAGTTGGTGTATTCAGCTAACCAACCTATCTCTGCTGATT 241  
 DB |||||  
 QY 231 TCGTACTAGTCCCGCTAAAGTTGGTGGATCCAGCCAACCACTTACCTTCGCGATT 290  
 DB |||||  
 QY 242 TCATTGGTGTCAATCTCCAGATTGAGCAATGTGATCCACTCTCTCATATGAGCAGGTC 301  
 DB |||||  
 QY 291 TCATTGGCATCACTCCAAATCCAGACCAAGTCATCGATTCCTTACCCGCCAGCGG 350  
 DB |||||  
 QY 302 TGAAGAAGCTTCTCTTCTTGGATCATCTCGATTTACCCATAAATTTGCTCTCAGCCAA 361  
 DB |||||  
 QY 351 TTAAGAAAATCTCTTCTCGGTTCTCTTGTATTATCCCAAAATTTGGCCGCAACCCA 410  
 DB |||||  
 QY 362 TTCCTGAGTCTGTTTGTAAACAGCATCGCTTGAACCAACTAATGATGCTATGTTG 421  
 DB |||||  
 QY 411 TCCCTGAAAACGCGCTCTTATCCGTCCTCCGAAACCCACCAACCAATGATGCGGTTG 470  
 DB |||||  
 QY 422 TTAAGATCGCTGGGATTAAGACTTGTAGGCTTATAGGATTCAGCAGCATGGATGCAA 481  
 DB |||||  
 QY 471 CCAAAATCGGGGATCAAAATGTCCAAAGCGTACAGATTCAGCAGCGCTTCGATGCTA 530  
 DB |||||  
 QY 482 TCCTTGGCATGCTACTAATCTCTATGGTCTTAATGCAATTTCCACCGGAGAAATCTC 541





Missouri, Columbia, MO 65211, USA  
 Location/Qualifiers  
 1. 1439  
 /organism="Zea mays"  
 /db\_xref="MaizeDB:638464"  
 /db\_xref="taxon:4577"  
 /clone="PC043177"  
 /clone\_lib="Maize Mapping Project/DuPont Cornsensus Library"  
 /notes="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"  
 BASE COUNT 297 a 357 c 443 g 342 t  
 ORIGIN  
 Query Match 43.6%; Score 408.2; DB 11; Length 1439;  
 Best Local Similarity 65.8%; Pred. No. 6e-109;  
 Matches 609; Conservative 0; Mismatches 313; Indels 3; Gaps 1;  
 QY 5 CTGCAAACTCTGCCAAATCTTCGTCGGGGTCATCGTGGTTTCGTTGGATCTGCCATTG 64  
 Db 1400 CCGAAGGACGCGAAGGTCTTCGTCGCGGCCACCGGGGCTCGTCGGCTCCGCCATCG 1341  
 QY 65 TCGCAAGCTTCAGGAACAAGTTTCACCAATCTCGTTCTTAAACACACAGCCGAGCTTG 124  
 Db 1340 TGGCGCGCTCCTCTCGCTAGGCTTCACCTCCGTCGTCGTCGCGACCCACGCCGAGCTG 1281  
 QY 125 ATCTCACTCGTCAAGCGCATGTTGAAATCCTTTCTTTTCTCAAGAAAGCCAGTTTATGTAA 184  
 Db 1280 ACCTGACTCGCAGGCGGACGTCGAGGCTTCTTCGCGGAGAACGCCCGCGCTACGCTG 1221  
 QY 185 TCCTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAACACACACCTATCCTGCTGATTTCA 244  
 Db 1220 TCCTCGCGCGCGCAAGGTGCGGGGAATCCACGCCAACTCCACCTTCCCGCGGACTTCA 1161  
 QY 245 TTGTGTCAATCTCCAGATTCAGACCAATGTGATCTCACTCTGCATATGAGCACGG---TG 301  
 Db 1160 TCGCGCGCACTCCAGATCCAGACCAACGTCGTCGAGCGCGCGCTGCGGCTCGGCTCGG 1101  
 QY 302 TGAAGAAGCTTCTCTCTTGGATCATCTCTGCAATTTACCCCTPAAATTTGCTCTCAGCCAA 361  
 Db 1100 TCGCAAGCTCTCTTCTCGGCTCTCTGCACTACCCCAAGTTCGCGCGCAGCCCA 1041  
 QY 362 TTCTGAGTCTGCTTTGTTAAACGATCGCTTGAACCAACTAATAGTGTGATGCTATTG 421  
 Db 1040 TCACGAGGGGCGCGCTCTCTTCGCGCGCGCTCGAGGCCACCAAGAGTGTGACCGCTG 981  
 QY 422 CTAAGATCGCTGGGATTAAGACTTGTTCAGGCTTATAGATTCACACGGATGGGATGCAA 481  
 Db 980 CCAAGATCGCGGGATCAAGATGTCCAGGCTACCGCATCCACACGGCTTCGACGCGG 921  
 QY 482 TCTCTGCATGCTTAAATCTCTATGGTCTCTAATGAAATTTCCACCCGGAGAAATTCCTC 541  
 Db 920 TCTCTGCATGCCCCACCAACCTGATGCGCGCACGACAACTTCCACCCGGAGAACTCGC 861  
 QY 542 ATGTGCTTCTGCTCTTATGAGGAGGTTTCACAGAGGCGAAAGTGAATGAGCGGAGGAG 601  
 Db 860 ACGTCTTCGCGCGCTCATCTCGTCGCTTCCACGAGGGCAAGGCTACTAAACGCCCCCGAGG 801  
 QY 602 TTGTGTGTGGGGTACAGGTAGTCCGTTGTAGGGAGTTCTTCATGTTGATGATTTGGCTG 661  
 Db 800 TCGTGTGTGGGGATCAGGCTCGCTCTGCGCGAGTTCCTGCACGTGATGATCTCGCGG 741  
 QY 662 ATGCTTGTGTTTCTTGTGTCGATCGATACAGCGGGTTGGAGCATGTTAAACATTCGAAAGT 721  
 Db 740 ACGCGGTCACTTCTGATGATCACTACTCCGCGATGGAGCATGTCAAATCTGGGGAGTG 681  
 QY 722 GTCAAGAAGTGACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGAGGTTGTGGTTTGAAG 781  
 Db 680 GGAGTGAGGTCACCATCAAGAGCTCGCCGAATAGTCAAGAGAGTGTGCGCTTCCAGG 621

QY 381 AACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATGCTAAGATCGCTGGGATTA 440  
 Db 181 TACTGACCCCTTAGAGCCACCAAAATGAATGGTATGCCATTGCCAAGATTGCTGGGATCA 240  
 QY 441 GACTGTCTAGGCTTATAGATTTCAGACGAGTGGGATGCAATCTCTGGCATGCTACTAA 500  
 Db 241 AATGTGCCAGGCTTACAGAAATTCAGCATAGTGGGATGCAATTCGGGAATGCCACCAA 300  
 QY 501 TCTCTATGTCCTTAATGACAATTTCCACCGGAGAATTTCTCATGTCTTCTCTGCTCTAT 560  
 Db 301 CTTATATGACCATATGACAAATTTTCATCCGGAATTCGATGTGTACCTGCTCTAT 360  
 QY 561 GAGGAGGTTCCAGAGCGGAAAGTGAATGAGCGGAGGAAGTTGTGTGGGTGCGGTACAGG 620  
 Db 361 GAGAAGGTTTCATGAGCGCAAGGTCAATGGTGCTAAGGAGGTGGTTGTGTGGGACCGG 420  
 QY 621 TAGTCCGTTGAGGAGTCTTCCATGTTGATGATTTGGCTGATGCTTGTGTTTCTTGCT 680  
 Db 421 AAGTCCATTTAGGAGGTTCTTGACGTTGACGATTTGGCAGACGCGGTTGTCTTCATGAT 480  
 QY 681 GGATCATACAGCGGTTGGAGCATGTTAACTTGAAGTGGTCAAGAGTGACTATTAG 740  
 Db 481 GGAAGATATAGCGGACTGGACATTTGAATGTAGGAGTGGAGGAGTTACTATTAA 540  
 QY 741 AGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGAAGCTTGGATGGGATG 800  
 Db 541 GGAATTTGGCTGAGTTGATGAAGAGAGTGGTGGGATTTGAAGGGATCTTGTGTTGGATTC 600  
 QY 801 CACTAAGCCAGATGGCAC 818  
 Db 601 TACTAAGCCTGATGGGAC 618

RESULT 8  
 AW041037  
 LOCUS EST283901 tomato mixed elicitor, BT1 Lycopersicon esculentum cDNA  
 DEFINITION clone cLET8A5, mRNA sequence.

ACCESSION AW041037  
 VERSION AW041037.1 GI:5899791  
 KEYWORDS EST.  
 SOURCE tomato.

ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.  
 1 (bases 1 to 652)  
 D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,  
 Roming,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,  
 Fraser,C.M., Venter,J.C., Martin,G.B., Tankale,S.D. and Giovannoni,  
 J.

REFERENCE Generation of ESTs from tomato leaf tissue  
 Unpublished (1999)

TITLE Contact: CUGI  
 JOURNAL Clemson University Genomics Institute  
 COMMENT 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

FEATURES Location/Qualifiers  
 source 1..652  
 /organism="Lycopersicon esculentum"  
 /cultivar="Rio Grande Ptor"  
 /db\_xref="taxon:4081"  
 /clone="cLET8A5"  
 /clone\_lib="tomato mixed elicitor, BT1"  
 /tissue\_type="leaf"  
 /dev\_stage="4-6 week old plants"  
 /lab\_host="XLI-Blue MRF"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; cLET - Inoculated with a variety of disease response  
 elicitors. Plants exposed to 2,6 dichloroisonicotinic

acid, BPH, jasmonic acid, ethylene, fenthion, BIX, okadaic acid, or systemin prior to tissue harvest. EcoRI site was destroyed during cloning."  
 BASE COUNT 179 a 158 c 118 g 197 t

Query Match 37.9%; Score 354.8; DB 10; Length 652;  
 Best Local Similarity 71.8%; Pred. No. 2.1e-93;  
 Matches 464; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 7 GACAAATCTGCCAAAATCTTCGTGCGGGTCATCGTGGTTGGTGGATTCGCCATTGTC 66  
 Db 7 GATACCTCCGTAAGATCTTCGTGCGCGCCACCGTGGACTCGTCGATCGCCGTGGTT 66  
 QY 67 CGCAAGCTTCAGGAACAAGGTTTCAACCAATCTCTGTTCTTAAACACACGCGAGCTTGAT 126  
 Db 67 CGGAACCTTTACCAATTAGGCTGCACAATCTCATCTCCGTACACATCCGATCCGAT 126  
 QY 127 CTCACCTCGTCAAGCGGATGTTGAATCTCTTCTTCAAGAGAACGCCAGTTTATGTAATC 186  
 Db 127 CTCACATCAATCCGCGTGCATCTCTTCGCGACGAGAAACCTCAATACGTCATC 186  
 QY 187 CTAGCAGCAGCTAAAGTTGGTGTATTACACCTTAACACACCTATCTCTGCTGATTTCAAT 246  
 Db 187 CTCGACGCGCGAAGTCGCGGCATACACGCAACAATACCTTACCCAGCTGATTTCAATC 246  
 QY 247 GGTGTCATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACACGCGTCTGAAG 306  
 Db 247 ACTATAATCTGCAATCCAAACGACGATCGTTTCATCTCTTCATCAACAAAGTTGAG 306  
 QY 307 AAGCTTCTCTCTTGGATCACTCTGCAATTTACCTAAATTTGCTCTCAACCAATTCCT 366  
 Db 307 AAGCTTCTCTCTTGGTTCTTCATGATTTACCTAAATTTGCTCTCAACCAATTCCT 366  
 QY 367 GAGTCTGCTTTGTTAAACAGCATCGCTTGAACCAACTAATGATGGTATGCTATTGCTAAG 426  
 Db 367 GAAAATGCATCTTTTAACTGCTCTTTGGAACCTCAAAATGAATGGTATGCAATTCGCAAA 426  
 QY 427 ATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGCGATGGATGCAATCTCT 486  
 Db 427 ATTGCTGGTATTAATGTGTCAAGCTTATAGATTGACGATTAATTTGATGCAATTTCA 486  
 QY 487 GGCATGCTACTAATCTCTATGTCCTTAATGACAATTTCCACCGGAGAAATTCATGATG 546  
 Db 487 GCAATGCTCAGAAATTTATATGTCAGAAATGCAAAATTTCCATCTCTGAGAAATTCATGAT 546  
 QY 547 CTTCTGCTCTTATGAGGAGGTTCCACGAGCGGAAAGTGAATGACGCGGAGAAATTTG 606  
 Db 547 TTGCTGCTGTTTGTACGTAGATTTTCATGAAGCAAAAGTTAAACAATCTTTGATAAAGGTT 606  
 QY 607 GTGTGGGTCACAGGTAGTCCGCTGAGGAGGTTCTTGCATGTTGATG 652  
 Db 607 GTGTGGGTCAGTCTGTTCTTTTAAAGGAAATTTTACATGTCGATG 652

RESULT 9  
 BH173899  
 LOCUS B216/R RFLP clone from soybean genomic DNA  
 DEFINITION clone B216 R primer, DNA sequence.

ACCESSION BH173899  
 VERSION BH173899.1 GI:15989266  
 KEYWORDS GSS.  
 SOURCE soybean.  
 ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 1 (bases 1 to 731)

REFERENCE Grant,D. and Shoemaker,R.C.  
 AUTHORS unpublished sequence of soybean RFLP probe  
 JOURNAL Unpublished (2001)

RESULT 10  
 BH173840  
 LOCUS  
 DEFINITION  
 A890/R RFLP clone from soybean genomic DNA linear GSS 09-OCT-2001  
 clone pA890 R primer, DNA sequence.  
 BH173840  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Glycine max  
 soybean.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.  
 REFERENCE  
 1 (bases 1 to 566)  
 Grant, D. and Shoemaker, R.C.  
 unpublished sequence of soybean RFLP probe  
 Unpublished (2001)  
 CONTACT  
 Contact: Grant D  
 Agronomy Department  
 USDA-ARS and Iowa State University  
 G304 Agronomy Hall, Ames, IA 50011-1010, USA  
 Tel: 515 294 1205  
 Fax: 515 294 2299  
 Email: dgrant@iastate.edu  
 single pass sequence  
 Class: RFLP clone.  
 FEATURES  
 source  
 1..566  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="pA890"  
 /clone\_lib="RFLP clone from soybean genomic DNA"  
 /note="Vector: pBS+; PstI-generated fragments of genomic  
 DNA. Library construction described by Keim, P. &  
 Shoemaker, R.C. (1988) Soybean Gen. News. 15:147-148."  
 BASE COUNT 151 a 119 c 143 g 153 t  
 ORIGIN  
 Query Match 37.6%; Score 352.4; DB 17; Length 566;  
 Best Local Similarity 76.5%; Pred. No. 1.1e-92;  
 Matches 432; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
 QY 197 CTAAGTTGGTGGTATTCACGCTAACACACCTACTCTGCTGATTTCATTGGTGTCAATC 256  
 DB 1 CCAAGTTGGTGGGATCCATGCCAACACACCTACTCTGCTGATTTCATTGGCCATCAACC 60  
 QY 257 TCCAGATTTCAGACCAATGTGATCCATCTGTCATATGACGCGGTGTAAGAGCTTCTCT 316  
 DB 61 TCCAAATCCAGACCAATGTGATTCGATTCGCAATGCGAATGGTGTGTAAGAACTGTGT 120  
 QY 317 TCCTTGGATCATCTCGGATTTACCTTAAATTTGCTCTCAGCCCAATTCCTGAGTGTGCTT 376  
 DB 121 TTTTGGGTTCCTCTTGCATTTTACCCCAATTTGACCCCAACCGGATTCGGAAGATGCTT 180  
 QY 377 TGTAAACAGCATCGCTTCAACCAATATGATGCTATGCTATTGCTTAAGATCGCTGGGA 436  
 DB 181 TGTCTACTGGACCTTAGAGCCCAATATGATGCTATGCTATGCCAAGATGCTGGGA 240  
 QY 437 TTAAGACTTGTGAGCTTATAGGATTCAGCACCGGATGGATGCAATCTCTGGCATGCTTA 496  
 DB 241 TCAAAATGTCCAGGCTTACAGATTCAGATAAGTGGGATGCAATTTCCGGAATGCCCA 300  
 QY 497 CTAATCTCTATGCTCTAATGACAAATTCACCCGGAGAAATTCATGCTGCTTCTGCTC 556  
 DB 301 CCAACTTATATGGACCATATGACAAATTTTCATCCCGAGAAATTCGATGCTGTACCTGCTC 360  
 QY 557 TTATCAGGAGGTTCACCGAGCGGAAGTGAATGACGCGGAGGAGTGTGCTGTGGGGA 616  
 DB 361 TCATGAGAAGGTTTCATGAGGCAAGGTCAATGTGCTTAAGGAGGTGGTGTGTGGGGA 420  
 QY 617 CAGGTAGTCCGTTGAGGAGGTTCTTGATGTGATGATTTGGCTGATGCTGTGTTTCT 676

Db 421 CCGAAGTCCATTGAGGAGTCTTGCAGCTGACGATTTGGCAGCGCGTGTCTTCA 480  
 QY 677 TGTGATCATACAGCGGTTGGAGCATGTTACATTTGGAAGTGTCAAGAGTACTA 736  
 Db 481 TGATGGAAGATATAGCGGATTCGAGCATTTGAATGTAGGAGTGGAAAGGAGTTACTA 540  
 QY 737 TTAGAGAGTTGGCTGAGTTGGTGAA 761  
 Db 541 TTAAGGAATTGGCTGAGTTGATGNA 565

RESULT 11  
 BQ116267 752 bp mRNA linear EST 22-JUL-2002  
 LOCUS EST601830 mixed potato tissues Solanum tuberosum cDNA clone STMDG80  
 DEFINITION 5' end, mRNA sequence.

ACCESSION BQ116267  
 VERSION BQ116267.2 GI:21917394  
 KEYWORDS EST.

SOURCE potato.  
 ORGANISM Solanum tuberosum

REFERENCE 1 (bases 1 to 752)  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,  
 Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and  
 Karamycheva, S.A.

TITLE Generation of a set of potato cDNA clones for microarray analyses  
 JOURNAL Unpublished (2002)  
 COMMENT On Apr 17, 2002 this sequence version replaced gi:20168216.

CONTACT: Robin Buell  
 The Institute for Genomic Research  
 9712 Medical Center Dr, Rockville, MD 20850, USA  
 Email: potato@igr.org  
 This clone is available through the Research Genetics, contact the  
 Research Genetics for further information 1-800-711-6195 or  
 cdna@resgen.com  
 Seq primer: T3.

FEATURES  
 source

Location/Qualifiers  
 1..752  
 /organism="Solanum tuberosum"  
 /cultivar="Kennebec or Binjite"  
 /db\_xref="taxon:4113"  
 /clone="STMDG80"  
 /clone\_lib="mixed potato tissues"  
 /tissue\_type="mixed tissues"  
 /lab\_host="SOLR"  
 /notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; supplier: Combination of untreated and Phytocphora  
 infestans-treated libraries of stolons, leaves, leaflets,  
 axillary buds of stem explants, petioles, germinating eyes  
 of tubers, or roots."

BASE COUNT 207 a 184 c 140 g 221 t

Query Match 37.5%; Score 350.6; DB 14; Length 752;  
 Best Local Similarity 69.9%; Pred. No. 4e-92;  
 Matches 501; Conservative 0; Mismatches 214; Indels 2; Gaps 2;

QY 7 GACAAATCTGCCAAATCTTCGTCGGGGTTCATCGTGGTTTGGTTCGATTCGCATTGTC 66  
 Db 35 GATACCTCCGCTAAGATCTTCGTCGGCGTCCCGTGGACTCGTCGGATCCGCGGTGTT 94

QY 67 CGGAAGCTTCAGGAACAAGTTTACCAATCTCGTTCTTAAACACACCGCGAGTTGAT 126  
 Db 95 CGGAACCTTTTACCAATTAGGTGCAAAATCTCTCTCCGTACACATTCGGATCTCGAT 154

QY 127 CTCACCTCGTCAAGCCGATGTGAATCTCTTTCTTCAAGAGAAGCCAGTTTATGATC 186  
 Db 155 CTCACCTAACCAATCCCGCGTGAATCTCTTTCGCGGACGAGAAACCTCAATACGTCATC 214

QY 187 CTAGCAGCAGCTAAAGTTGGTGGTATTACGCTAAACAACACCTATCTCTGCTGATTTCATT 246  
 Db 215 CTCGCGCGCGAGGTGCGCGGCATACACGCAACAATACTTACCAGCTGATTTCATC 274  
 QY 247 GGTGCAATCTCAGATTTCAGACCAATGTGATCCACTCTGCAATATGAGACCGGTGGAAG 306  
 Db 275 ACTATAAATCTTCAAAATCCAAACAAACGTCATCGTTTCATCTCTCAATCACAAAGTTCAG 334  
 QY 307 AAGCTTCTCTCTTGGATCATCTTGCATTTTACCCCTAAATTTGCTCCCTCAGCAATTCCT 366  
 Db 335 AAGCTTCTGTTCCITGGTTCTTTCATGTATTTTACCCTAAATTTGCTCCCTCAACCAATTCCT 394  
 QY 367 GAGTCTGCTTTTAAACAGCATCGCTTGAACCAACAATTAATGAGTGGTATCTATTGCTAAG 426  
 Db 395 GAAATGACATTTTAACTGCCCTTTGGAACCTACAATTAATGATGTATGCAATTCGAAA 454  
 QY 427 ATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGGAATGGGATCAATCTCT 486  
 Db 455 ATTGCTGGTATCAAAATGTGTCAAGCTTATAGATTGCAGCATAACTTTGATGCAATTTCA 514  
 QY 487 GGCATGCTCTACTAATCTCTATGGTCTCTAATGACAATTTCCACCGGAGAAATTTCTCATGTG 546  
 Db 515 GCAATGCCACGAAATTTATATGGTACGAATGACCAATTTCCATCTCTGAGAAATTTCTATGTT 574  
 QY 547 CTTCCTGCTCTTATCAGGAGGTTCCACGAGGCGAAAGTGAATGGAGCGGAGGAAGTTGTG 606  
 Db 575 TTGCTCTGCTTTGTTACGTAGATTTTCATGAAGCAAAAGTAAACAATCTTGATAAGTTGGTT 634  
 QY 607 GTGTGGGG-TACAGTAGTCTCGTTGAGGAGTCTTTCGATGTGATGATTTGGCTGATGC 665  
 Db 635 GTGTGGGAAACTGCTTCTCTTTAAGGAAATTTTACATGTGCGATGATTTAACCGATGC 694  
 QY 666 TTGCTGTTTCTTGTGGATCGATCAGCGGGTTGGAGCATGTTAACATTTGGAAGTGG 722  
 Db 695 ACTAGTGTTTTGGTGGAGAAATTATAGTGAATTTAGA-CATGTTTATGTTGGGAGTGG 750

RESULT 12  
 BQ440809

LOCUS GA\_Ea0010D16f Gossypium arboreum 7-10 dpa fiber library Gossypium

DEFINITION GA\_Ea0010D16f, mRNA sequence.

ACCESSION BQ440809

VERSION BQ440809.1 GI:13350461

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

REFERENCE 1 (bases 1 to 687)

AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

TITLE An integrated analysis of the genetics, development, and evolution

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGGG  
 High quality sequence stop: 683.

FEATURES  
 source

Location/Qualifiers  
 1..687  
 /organism="Gossypium arboreum"  
 /strain="AKA"  
 /cultivar="8400"  
 /db\_xref="taxon:29729"  
 /clone="GA\_Ea0010D16f"  
 /clone\_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue\_type="Fibers isolated from bolls harvested 7-10  
 dpa"  
 /lab\_host="E. coli"  
 /note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 167 a 213 c 134 g 173 t

Query Match 36.3%; Score 339.8; DB 12; Length 687;  
 Best Local Similarity 72.5%; Pred. No. 5.7e-89;  
 Matches 440; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 2 TGTCTGACAAATCTGCCAAATCTTCGTCGGGTTCATCGTGGTTGGTTGGATCTGCCA 61  
 DB 81 TCTCGAACAAATCAGCGAAGATCTTCTAGAGCCCTCGAGCCTAGTTGGTTCTGCCA 140  
 QY 62 TTGTCCGCAAGCTTCAGGAACAAAGGTTTCCACCAATCTCGTTCTTAAACACACGCGGAGC 121  
 DB 141 TAGTTTGGTAACTTCAGTCCCTAGGCTTTTACCAACCTCTACTCCGACCCATGCCGATC 200  
 QY 122 TTGATCTCACTGTCAAGCGGATGTAATCTCTTTCTTCAAGAGAACGCGTTTATG 181  
 DB 201 TGGACCTTATCGCCAAATCCGAGTCGAATCTCTTTCGCGCATGAGAAACCTCACTATG 260  
 QY 182 TAATCTACGACGCTAAAGTTGGTGTATTACGCTTAACACACCTATCTGCTGATT 241  
 DB 261 TGTACTAGTCCGCTAAAGTTGGTGGATCCACGCCAACACACTTACCTCTGCCGATT 320  
 QY 242 TCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACGCGTG 301  
 DB 321 TCATTGCCATCAACCTCCAAATCCAGACCAACGTATGATCTCTTACCGCCACGGG 380  
 QY 302 TGAAGAAGCTTCTCTCTTGGATCATCTCTGATTTACCTAAATTTGCTCTCAGCCAA 361  
 DB 381 TTAAGAAAAATCTCTCTCTCGTTCTCTTGTATTATCCCAAATTTGGCGCGCAACCA 440  
 QY 362 TTCCCTGAGTCTGTTTAAACAGCATCGCTTGAACCAACTAATGATGATGCTATTG 421  
 DB 441 TCCCTGAACACGCGCTCTTATCCGTCCTCGTGAACCCACCAACGAATGGTATGCGGTTG 500  
 QY 422 CTAAGATCCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACCGGATGGATGCA 481  
 DB 501 CCBAATCCGGGGATCAAAATGTGCCAAGCGTACAGATTCAGACGCTTCGATGCTA 560  
 QY 482 TCTTGGCATGCTACTAATCTATGTCCTAATGACAATTTCCACCGGAGAAATCTC 541  
 DB 561 TTTCCGCTATGCGCAAAATTTGTACGCGCCCAACGATTAATTTCCATCCGGAATAATCC 620  
 QY 542 ATGTGCTTCTGCTCTTATGAGGAGGTTCCACGAGCGGAAGTGAATGAGCGGAGAG 601  
 DB 621 ACGTTTTCGCGCTTTGATGCGGAGGTTCCATAAGCCAAAGTGAAGGAGCAAGAAG 680  
 QY 602 TTGTGGT 608  
 DB 681 TAGTTGT 687

RESULT 13  
 AI897511  
 LOCUS 614 bp mRNA linear EST 18-MAY-2001  
 DEFINITION EST266954 tomato ovary, TAMU Lycopersicon esculentum cDNA clone  
 CLE227F14, mRNA sequence.

ACCESSION AI897511  
 VERSION AI897511.1  
 KEYWORDS GI:5603413  
 EST.

SOURCE tomato.

ORGANISM

Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;  
 Lycopersicon.

1 (bases 1 to 614)

REFERENCE  
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.  
 , Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman

C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley  
 ,S.D. and Giovannoni,J.  
 Generation of ESTs from tomato carpel tissue  
 Unpublished (1999)  
 Contact: CUGI  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 5 prime sequence.

FEATURES  
 Location/Qualifiers  
 1..614  
 /organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="CLE227F14"  
 /clone\_lib="tomato ovary, TAMU"  
 /tissue\_type="carpel"  
 /dev\_stage="5 days pre-anthesis to 5 days post-anthesis"  
 /lab\_host="XLI-Blue MRP"  
 /note="Vector: pBlueScript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; cLED - Tomato Carpel EST Library. OligoDT-primed and  
 directionally cloned cDNA in vector Lambda ZAP II with 5'  
 and 3' ends located at the EcoRI and XhoI sites,  
 respectively."  
 BASE COUNT 168 a 150 c 107 g 189 t  
 ORIGIN

Query Match 36.2%; Score 338.6; DB 9; Length 614;  
 Best Local Similarity 72.2%; Pred. No. 1.2e-88;  
 Matches 440; Conservative 0; Mismatches 169; Indels 0; Gaps 0;  
 QY 13 TCTGCCAAATCTTCGTCGGGTTCATCGTGGTTGGTTGGATCTGCCATTTGCCGAAG 72  
 DB 5 TCGCTAAGATCTTCGTCGTTGGCCACCGTGGACTCGTCGGATCCGCGGTTCGAAA 64  
 QY 73 CTTTCCAGAACAGGTTTACCAATCTCGTTCTTAAACACACGCCGAGCTTGATCTCACT 132  
 DB 65 CTTTACCAATTAGGCTGCACAAATCTCATCTCCGTCACACATTCGATCTCGATCTCACT 124  
 QY 133 CGTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTATGTAATCTTAGCA 192  
 DB 125 AATCAATCCGCGTGAATCTCTTCGCGGAGAGAAACCTCAATACGTCATCTTCGCA 184  
 QY 193 GCAGCTAAAGTTGGTGGTATTTCACGCTAAACACACCTATCTCTGCTGATTTTCATTTGGTGC 252  
 DB 185 GCGCGAAGTCCGCGCATACACGCAACATCTTACCAGCTGATTTTCATCACTATA 244  
 QY 253 AATCTCCAGATTACAGACCAATGTGATCCACTCTGCTATAGAGCAGCGTGAAGAGCTT 312  
 DB 245 AATCTGCAATCCAAACGAAACGTTATCTTCATCTTCAATCACAAGTTCAGAAGCTT 304  
 QY 313 CTTCTCCCTGGATCATCTTCGATTTACCTTAATTTGCTCTCCAGCAATTCCTGAGTCT 372  
 DB 305 CTGTTCTCTGGTTCTTCATGTATTATACCTTAATTTGCTCTCCCAACCAATTCCTGAAAT 364  
 QY 373 GCTTTGTTAACAGATCGCTTGAACCAACTAATAGTGGTATGCTATTGTTAAGATCGCT 432  
 DB 365 GCATTTTAACTGCTCTTTGGAACTACAAATGATGATGCAATTCGGAATTTGCT 424  
 QY 433 GGGATTAAAGCTTTGTCAGGCTTATAGGATTCAGACGGAATGGGATGCAATCTCTGGCATG 492  
 DB 425 GGTATTTAAATGTGTCAAGCTTATAGATTGCAGCATAACTTTTATGATCAATTTTCAGCAATG 484  
 QY 493 CCTACTAATCTTATGGTCTTAATGACAAATTTCCACCGGAGAAATTCATGCTCTCTCT 552  
 DB 485 CCTACGAATTTATGTTAGTACGAATGCAATTTCCATCTCTGAGAAATTTCTCATGTTTGCCT 544  
 QY 553 GCTCTTATGAGGAGGTTCCACGAGGCGAAAGTGAATGGAGCGGAGGAAGTTTGGTGTGG 612  
 DB 545 GCTTTGTTACGTAGATTTCATGAGCAAAAGTTAACAATCTTGATAAAGTTTGTGTGG 604  
 QY 613 GGTACAGGT 621

```
Db 605 GGTACTGGT 613
|||||
A1726948 664 bp mRNA linear EST 11-JUN-1999
LOCUS BNLH16903 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
DEFINITION to (AF045286) GDP-4-keto-6-deoxy-D-mannose-3
5-epimerase-4-reductase [Arabidopsis thaliana], mRNA sequence.
ACCESSION A1726948
VERSION A1726948.1 GI:5045800
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 664)
Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
ESTs from developing cotton fiber
Unpublished (1999)
Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burrb@bnl.bnl.gov
Seq primer: T3 Primer.
Location/Qualifiers
1. .664
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/clone_lib="Six-day Cotton fiber"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"
/notes="Vector: pBluescript II KS-+"
BASE COUNT 161 a 204 c 128 g 169 t 2 others
ORIGIN
Query Match 34.8%; Score 326; DB 9; Length 664;
Best Local Similarity 71.7%; Pred. No. 6.6e-85;
Matches 439; Conservative 0; Mismatches 172; Indels 1; Gaps 1;
QY 2 TGCTGACAAATCTGCCAAATCTTCGTCGGGTCATCGTGGTTGGTTGGATCTGCCA 61
Db 54 TCTCGAACAAATCAGCGAAGATCTTCGTANCAAGCCATCGAGGCTAGTTGGTTCTGCCA 113
QY 62 TTCTCGCAAGCTTCAGGAACAGGTTTCCACCAATCTCGTTCTTAAACACACGCGAGC 121
Db 114 TANTTCGTAACTTCAGTCCCTAGGCTTTACCAACCTCTCTACGCAACCACTTACCCCTGCCGATC 173
QY 122 TTGATCTCACTCGTCAAGCGGATGTGCAATCTCTTTCTCAAGAGAAGCCAGTTTATG 181
Db 174 TGGACCTTACTCGCAATCGAGCTCGAATCTCTTTCGCCGATGAGAACCCTCACTAGC 233
QY 182 TAATCTCAGCAGCTAAAGTTGGTGGTATTACGCTAACACACCTATCTCTGCTGAT 241
Db 234 TCGTACTAGTCGCGCTAAAGTTGGTGGATCCACGCCAACCACTTACCCCTGCCGAT 293
QY 242 TCATTGGTGCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCGG 301
Db 294 TCATTGCCATCAACCTCCAAATCCAGACCAAGTCATGATTCTCTTACCGCCAGCGG 353
QY 302 TGAAGAAGCTTCTCTCTTGGATCATCTGCAATTTACCTTAAATTTGCTCTCAGCCAA 361
Db 354 TTAACAAATCTCTCTCTCGGTTCTCTTGATTATTCACCAAGTTTGGCGCGCAACCA 413
QY 362 TTCTGAGTCTGCTTGTGTTAAACAGCATCGTTGTAACCAACTAATGATGCTGATTTG 421
|||||
```

```
Db 414 TCCCTGAAATCGCTCTTTATCCGGTCCCTCGAACCACCAACGAATGGTATGCCGTTG 473
QY 422 CTAAGATCGCTGGGATTAGACTTTGTCAGGCTTTATAGGATTCAGACCGATGGATGCAA 481
Db 474 CCAAGATCGGGGATCAAAATGTGCCAAGCGTACAGAATTCAGCACGGCTTCGATGCTA 533
QY 482 TCTCTGGCATGCCTACTATCTATCTATGGTCTTAATGACAAATTTCCACCGGAGAAATCTC 541
Db 534 TTTCCGCTATGCGGCAAAATTTGTACGGCCCCCAACGATTAATTTCCATCTCTGAAATTTCC 593
QY 542 ATGTGCTTCTCTCTCTTATGAGGAGTTCCACAGCGGAAAGTGAATGAGCGGAGGAAG 601
Db 594 ACGTTTTCGCCGCTTTGATGCGGAGGTTCC-TAAAGCCAAAGTGGATAGACGCAAAAAAAG 652
QY 602 TTGTGGTGGG 613
Db 653 TAGTTGTGGG 664

RESULT 15
A1487671 590 bp mRNA linear EST 18-MAY-2001
LOCUS EST245993 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION CLED13M21, mRNA sequence.
ACCESSION A1487671
VERSION A1487671.1 GI:4383042
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 590)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
,S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1. .590
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLED13M21"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 166 a 151 c 96 g 177 t
ORIGIN
Query Match 34.4%; Score 322.2; DB 9; Length 590;
Best Local Similarity 72.5%; Pred. No. 8.1e-84;
Matches 417; Conservative 0; Mismatches 158; Indels 0; Gaps 0;
QY 13 TCTGCCAAATCTTCGTCGGGTCATCGTGGTTGGTGGATCTGCCATCTCGCGCAAG 72
Db 5 TCCGCTAGATCTTCGTCGGGCCACCGTGGACTCGTCGGATCCGCGTGTTCGGAA 64
QY 73 CTTTAGGAACAGGTTTCCCAATCTCGTTCTTAAACACACGCGAGCTTGATCTCACT 132
|||||
```

Db 65 CTTTACCAATTAGGTCGACAAATCTCATCTCCGTCACATTCGGATCTCGATCTCACT 124  
QY 133 CGTCAAGCCGATGTTGAATCCTTCTTTCTCAAGAGAGCCAGTTTATGTAATCCTAGCA 192  
Db 125 AATCAATCCGCGTCGAATCCTTCTGCGGACGAGAAACCTCAATACGTCATCCTCGCA 184  
QY 193 GCAGCTAAAGTTGGTGTATTCAAGCTACACACCTATCCTGCTGATTTGATGTC 252  
Db 185 GCGCGAAAGTCGGCGGCATACACGCAAACTACTTACCCAGCTGATTTCACTATA 244  
QY 253 AATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGAGCAGGTCGGAAGCTT 312  
Db 245 AATCTGCAATCCAAACGACGTTATCGTTTCATCCTTCAATCACAAGTTCAGAGCTT 304  
QY 313 CTCTTCCTTTGGATCCTCTGCAATTTACCTAAATTTGCTCCTCAGCCAAATCCTGATCT 372  
Db 305 CTGTTCTTGGTCTTTCATGTATTTACCTTAAATTTGCTCCTCAACCAATTCCTGAAAT 364  
QY 373 GCTTTGTTAACAGCATCGCTTGAACCACTAATGATGGTGTATGCTATGCTAAGATCGCT 432  
Db 365 GCACCTTTAACTGCTCCTTTGGAACCTACAAATGAATGGTATGCAATTCGAAATTTGCT 424  
QY 433 GGGATTAAAGACTTGTCAAGCTTATAGGATTCAGACGGATGGATGCAATCTTGGCATG 492  
Db 425 GGTATTAAATGTGTCAGCTTATAGATTGCGACATACTTTGATGCAATTTCAGCAATG 484  
QY 493 CCTACTAATCTCTATGGTCTTAATGACAAATTTCCACCGGAGAAATTCATGTGCTTCT 552  
Db 485 CCTACGAATTTATATGGTACGAATGACAATTTCCATCCTCGAGAAATTCATGTTTGCCT 544  
QY 553 GCTCTTATGAGGAGTTCCACGAGGGAAGTAA 587  
Db 545 GCTTTGTTACGTAAATTTTCATGAACAAAAGTTAA 579

Search completed: June 3, 2003, 04:00:13  
Job time : 1678 secs







QY 306 GAAGCTTCTCTTGGATCATCTGCAATTAACCTAAATTTGCTCCTCAGCAATTC 365  
 Db 1704549 GCGGCTCTGTTCTGGGCTCGTGTGATCTACCGAACTCCGCCCGCAGCGATCCC 1704608  
 QY 366 TGAGTCTGCTTTGTTAAACAGCATCGCTTGAACCACTAATGAGTGTATGCTATTGCTAA 425  
 Db 1704609 GGAAGCGCGCTGCTACCGGCTCGTTGGAGCGACCAAGCAGCGTACGCGATCGCCAA 1704668  
 QY 426 GATCGCTGGGATTAAGACTTTGTGAGGCTTATAGGATTCAGCAGCGATGGGATGCAATCTC 485  
 Db 1704669 AATCGCGGATTTCTGGGCTCAGGCGTGGCCGCAACATGGCTCGCGTGGATCTC 1704728  
 QY 486 TGGATGCTCTAATCTTATGCTCTAATGCAATTTCCACCGGAGAAATTTCTCATGT 545  
 Db 1704729 GGGGATGCCCAACCACTGTACGGGCGAGCGCACTTTTCGCGCTCGCATCT 1704788  
 QY 546 GCTTCTGCTCTTATGAGGAGTTTCACGAGGCGAAAGTGAATGGAGCGGAGGATGT 605  
 Db 1704789 GCTGCGGCACTCATCGCGCTATGACGAGGCGCAAGCCAGTGGCGCGCCCAACGTGAC 1704848  
 QY 606 GGTGTGGGTPACAGGTAGTCCGTTGAGGAGGATTTCTGCAATTTGATGATTTGGCTGATGC 665  
 Db 1704849 CAATGGGCGACCGGACCGCCCGAGGAGTTGCTGCACTGACGACCTGGCGAGCGC 1704908  
 QY 666 TTGTGTTTCTGCTGATCGATACACGCGGTTGGAGCATGTTAACTTGGAAAGTGTCA 725  
 Db 1704909 ATGCTGTATCTGCTGCAATTTTCAGCGGCGGACCCATGTCAACGTGGAAACCGCAT 1704968  
 QY 726 AGAAGTGAATTTAGAGAGTTGCTGAGTGGTGAAGAGTTGTTGGTTTGAAGGAA 785  
 Db 1704969 CGACCAACATCGGCGAGATCCCGGAGATGGTGGCTCGCGGCTAGGCTATAGCGGCA 1705028  
 QY 786 GCTTGGATGGATTTGCACTAAGCCAGATGGCAGACCGAGGAACTTTATGACAGCTCAAA 845  
 Db 1705029 AACCCGCTGGATCAAGCAACCGGACGAAACACACGCAACTGCTGGATTTCCGT 1705088  
 QY 846 GCTCGGCTTTGGGTTGGACACTAAGTTTCTTTAGAGATGGTCTGAGCCAAACTTA 905  
 Db 1705089 GCTACGGGAGCGGATGGCGCTTCGATCGGCTCGCGACGCGATCGAGGCGAGT 1705148  
 QY 906 TGATTGGTATTTCAAGATG 925  
 Db 1705149 GCGGTGTATCGGAGCAG 1705168

RESULT 2

US-09-060-756-1  
 ; Sequence 1, Application US/09060756  
 ; Patent No. 6183957  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cole, Stewart  
 ; APPLICANT: Buchrieser-Brosch, Roland  
 ; APPLICANT: Gordon, Stephen  
 ; APPLICANT: Billault, Alain  
 ; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
 ; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
 ; FILE REFERENCE: 3495-0169  
 ; CURRENT APPLICATION NUMBER: US/09/060,756  
 ; CURRENT FILING DATE: 1998-04-16  
 ; NUMBER OF SEQ ID NOS: 743  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 12732  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 US-09-060-756-1

Query Match 29.4%; Score 275.2; DB 4; Length 12732;  
 Best Local Similarity 56.2%; Pred. No. 1.1e-82;  
 Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;  
 QY 6 TGCAAAATCTGCCAAATCTTCGTCGGGTGTCATCGTGGTTGGATCTGCCATTTGT 65

Db 8106 TGACCGCGCGCGCGGCTTACATCGCGGGATCGCGCCTGGTCCGGTCCGCGTGTCT 8165  
 QY 66 CGCAAGCTTTACAGAACAAAGGTTTACCAATCTCGTTCTTAAACACACACGCCGAGCTTGA 125  
 Db 8166 AGCAGCTTTGCGGGCGCGGGTTTACCAACTCTGCTGGTTCGGTTCAGCGCCGAGCTTGA 8225  
 QY 126 TCTCACTCTCAAGCCGATGTTTGAATCTCTTTCTCAAGAGAAAGCAGTTTATGTAAT 185  
 Db 8226 TCTGACGGATCGGGCGCGAGCTTCGACTTCGTTCTCGAGTCGAGSCCGCAGTCTCAT 8285  
 QY 186 CTTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTTAACAACACCTATCTCTGCTGATTTTCAT 245  
 Db 8286 CGACGCGCGCGCGCGGCTCGGGGATCTCTGGCCAAACGACACCTACCGCGGATTTCTCT 8345  
 QY 246 TGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATAGCACGCGTGTGA 305  
 Db 8346 GTCGGAAACCTCCAGATCCAGGTCAACCTGTGATCGCGCTGGCGCGCGGGTGC 8405  
 QY 306 GAAGCTTCTCTTCTTGGATCATCTGCAATTTACCTAAATTTGCTCCTCAGCAATTC 365  
 Db 8406 GCGGCTGCTGTTCTCGGCTCGTGTGATCTACCGAACTCGCCCGCAGCGATCCC 8465  
 QY 366 TGAGTCTGCTTTGTTAAACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATTGCTAA 425  
 Db 8466 GGAGAGCGCGCTGCTACCGTCCGTTGGAGCGGACCAACGACGCGTACGCGATCGCCAA 8525  
 QY 426 GATCGTGGGATTAAGACTTTGTGAGGCTTATAGGATTCAGCAGCGATGGGATGCAATCTC 485  
 Db 8526 AATCGCGGATCTCTTGGCTCCAGCGGTCGCGCGCAACATGGCTCGCTCGGCTGATCTC 8585  
 QY 486 TGCGATGCTCTAATCTCTATGCTCTAATGACAAATTTCCACCGGAGAAATTTCTCATGT 545  
 Db 8586 GCGGATGCCCAACCTGTACGCGGCGAGCGCAACTTTTCGCGCTCGGCTCGCATCT 8645  
 QY 546 GCTTCTGCTCTTTATGAGAGGTTTCAACGAGCGAAAGTGAATGAGAGCGGAGAAAGTTGT 605  
 Db 8646 GCTGCGGCACTCATCCGCGCTATGACGAGGCGCAAGCAGTGGCGCGCCCAACGTGAC 8705  
 QY 606 GGTGTGGGTACAGGTAGTTCGTTGAGGAGTTTCTTGCATGTTGATGATTTGCTGATGC 665  
 Db 8706 CAACTGGGCGACCGGCAACCGCCGCGAGGAGTTGCTGACGTCGACGACTGGCGAGCGC 8765  
 QY 666 TTGTGTTTTTCTGCTGATCGATACAGCGGTTGGAGCATGTTAACTTGGAAAGTGGTCA 725  
 Db 8766 ATGCTGTATCTGCTGGAACATTTTCGCGGCGGACCCCATGTCAGTGGGAAACCGCAT 8825  
 QY 726 AGAAGTGAATTTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTTGAAGGGAA 785  
 Db 8826 CGACCAACACCATCGCGGAGATCGCGGATGTTGCTCGCGGTAGGCTATAGCGGCGA 8885  
 QY 786 GCTTGGATGGGATTCACCTAAGCCAGATGGCAGACCGAGGAACTTATGACAGCTCAAA 845  
 Db 8886 AACCCGCTGGGATCCAGCAACCGGACGAAACACACGCAACTGCTGGATGTTTCGT 8945  
 QY 846 GCTCGCGCTTTGGGTTGGACACCTAAGGTTTTCTTTAGAGATGGTCTGAGCCAAACTTA 905  
 Db 8946 GCTACGGGAGCGGATGGCGGCTTCGATCGCTGCGCGAGCGCATCGAGCGGCGGT 9005  
 QY 906 TGATTGGTATTTGAAGATG 925  
 Db 9006 GCGGTGTATCGGAGCAG 9025

RESULT 3

US-09-103-840A-1  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.

```

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      29.4%; Score 275.2; DB 4; Length 4411529;
Best Local Similarity 56.2%; Pred. No. 5e-81;
Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

Qy   6 TGACAAATCTGCNAAAATCTTCGTCGGGGTCAATCGFHGGTTTGTTGGTAGACTCTGCCATTGT 65
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704120 TGACCGCGCGCCGGGTCTACATCGCCGGCATCGCGGCTGCTGTCGGGTCCGCGTGTCT 1704179

Qy   66 CCGCAGAGCTTCAGGAACAAGGTTTTCACCAATCTCGTTCTTTAAAAACAACGCCCGAGCTTGA 125
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704180 ACGCAGTTTTCGGGCGCGGGGTTCACCAACCTGCTGTTGGGTTCACGCCCGGAGTTGA 1704239

Qy   126 TCTCACTCGTCAAGCCGATGTGAATCCITCTTTTCTCAAGAGAAGCCAGTTTATGTAAT 185
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704240 TCTGACGGATCGGGCGCGAGCTTCGACTTCGTTCTCGAGTCGAGGCGCGAGTGTCTCAT 1704299

Qy   186 CTTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAACCAACCACTATCTCTGCTGATTTCA 245
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704300 CGACCGCGCGCCGGGTCCGGCGCATCTCTGGGCCAACACCACTACCCGCCCGGATTTCT 1704359

Qy   246 TGGTGTCAAATCTCCAGATTCCAGACCATTGATCCACTCTGCAATATAGCACACCGTGTGAA 305
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704360 GTCCGAAAACCTCCAGATCCAGGTCAACCTGCTGGATGCGCCGCTGGCGCGCGGGTGCC 1704419

Qy   306 GAAGCTTCTCTCTCTGGATCATCTGTCATTTACCCCTAAATTTGCTCTCTCAGACCAATTCC 365
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704420 GCGGCTGCTGTTCTTGGGCTCGTGGTGCATCTACCCGAACTCGCCCGCAGACCGATCCC 1704479

Qy   366 TGAGTCTGCTTTGTTAAACAGCATCGCTTGTAACCAAATAAGTGGTGTATGCTATTGCTAA 425
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704480 GGAGAGCGCGCTGCTCACCGGTCCGTTGGAGCCGACCAACGACGCTACGCGATCGCAA 1704539

Qy   426 GATCGCTGGGAATTAAGACTTGTGAGGCTTATAGATTTCAGACGGATGGGATGCAATCTC 485
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704540 AAATCGCGCGCATCTCTGCGGTCCAGGCGGTGCGCGCCGCAACATGGCGTCCGCTGGATCTC 1704599

Qy   486 TGGCATGCCCTACTAATCTCTATGTCCTAATGACAAATTTCCACCGGAGAAATTTCTCATGT 545
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704600 GGGATGCCCAACCACTGTACGGCCAGGGCAACATTTTCGCGCTCGCGCTCGCATCT 1704659

Qy   546 GCTTCTCTGCTCTTATGAGGAGTTTCCACGAGGCGGAAAGTGAATGGAGCGGAGGAAGTTGT 605
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704660 GCTCGCGCACTCATCCGCGCTATGACGAGGCCAAGACCCAGTGGCGGCCCCAACGTCAC 1704719

Qy   606 GGTGTGGGTACAGGTAGTCGGTTGAGGAGATTCCTTCATGTTGTAGATTGTTGCTGATGC 665
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704720 CAATCGGGGACCGGCACGCCGCCGACGGGAGTTGTCGACGTCGACACCTGGCGAGCGC 1704779

Qy   666 TTGTGTTTTCTGCTGGATCGATACACGCGGTTGGAGCATGTTAAACATTGGAAGTGTCA 725
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704780 ATGCTGTATCTGCTGGAAATTTTGAACGGGCGCA CCCATGTCAACGTGGGAACCGGCAT 1704839

Qy   726 AGAAGTGACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGAGTTGTTGGTTTTGAAGGGAA 785
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704840 CGACCAACCACTCGGCGAGATCGCCGAGATGGTCCGCTCGCGGTAGGCTATAGCGCGA 1704899

Qy   786 GCTTGGATGGGATTGCACTAAGCAGATGGCAACCGAGGAAACTTATGACAGCTCAAA 845
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1704900 AACCCGCTGGGATCCAAAGCAAAACCGACCGGAAACACACGCAAACTGCTGGAATGTTCCGT 1704959

```

Qy	846	GCTCGCGTCTTTGGTTGGACACTAAGTTTCTCTTAGAGATGGTCTGAGCCAACTTA	905
Db	1704960	GCTACGGAGCGCGATGCGGCGCTTCGATCGCGCTCGCGGACGGCATCGAGGCGACGGT	1705019
Qy	906	TGATTGCTATTGAGAAATG	925
Db	1705020	GCGTGGTATCGCGAGCAGC	1705039
RESULT 4			
US-09-453-702B-121/c			
: Sequence 121, Application US/09453702B			
: Patent No. 6365723			
GENERAL INFORMATION:			
APPLICANT: Blattner, Frederick R.			
Burland, Valerie			
Perna, Nicole T.			
Plunkett, Guy			
Welch, Rod			
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157			
NUMBER OF SEQUENCES: 265			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Quarles & Brady			
STREET: 1 South Pinckney Street			
CITY: Madison			
STATE: WI			
COUNTRY: US			
ZIP: 53701-2113			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Word Perfect 8.0			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/09/453,702B			
FILING DATE: 03-Dec-1999			
CLASSIFICATION: <Unknown>			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: 60/110,955			
FILING DATE: 04-DEC-1998			
ATTORNEY/AGENT INFORMATION:			
NAME: Seay, Nicholas J.			
REGISTRATION NUMBER: 27386			
REFERENCE/DOCKET NUMBER: 960296.95017			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (608) 251-5000			
TELEFAX: (608) 251-9166			
INFORMATION FOR SEQ ID NO: 121:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 14187			
TYPE: nucleic acid			
STRANDEDNESS: double			
TOPOLOGY: linear			
MOLECULE TYPE: DNA (genomic)			
SEQUENCE DESCRIPTION: SEQ ID NO: 121:			
US-09-453-702B-121			

Query Match	28.8%	Score 269.8	DB 4	Length 14187
Best Local Similarity	57.1%	Mid. No. 8.3e-81		
Matches 530	Conservative	0	Mismatches 372	Indels 27
Gaps	1			
22	ATCTTCGTCGCGGTCATCGTGGTTTGGTTCGATCTGCCATTGTCGCAAGCTTCAGGAA	81		
6374	ATTTTATTTGCTGGTGCACCAAGGAATGGTTGGATCAGCTATTACCCGACGCCTCAAAACA	6315		
82	CAAGGTTTCACCAATCTCGTTCTTAAAAACAACGCGGAGCTTGTATCTCACTCGTCAAGCC	141		
6314	CGTGATGATGTTGAGTTGGTTTACGTACTCGGATGNAATTGACTGTTGGATAGTAGC	6255		
142	GATGTTGAATCCTCTTTCTCAAGAGAAGCCAGTGTATGTAATCTCTAGCAGCAGCTAAA	201		
6254	GCTGTTTTGGATTTTTTTCTTCACAGAAATCAACACAGGTTTATTTTGGCAGCAGCAAA	6195		



APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: ANTIGENS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/937,972  
FILING DATE: Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PP-0400 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1352 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BLADNOT04  
CLONE: 1318190  
US-08-937-972-4

Query Match 11.0%; Score 103.2; DB 2; Length 1352;  
Best Local Similarity 48.9%; Pred. No. 8.4e-25;  
Matches 341; Conservative 0; Mismatches 348; Indels 9; Gaps 2;

QY 124 GATCTCACTCGTCAAGCCGATGTTGAATCCTTCTTCTCAAGAGAGCCAGTTTATGTA 183  
DB 223 GATCTCAGGATACAGACAGACCCGCCCTGTTTGAGAGGTCCAACCCACACAGTC 282  
QY 184 ATCTAGACAGCACTAAAGTTGGTGTATTACGCTTAACACACCTATCTCTGCTGATTC 243  
DB 283 ATCCATCTTGTGCAATGGTGGGGCCCTGTTCCGGAATATCAATACAAATTTGGACTTC 342  
QY 244 ATGGTGTCAATCTCAGATTAGACCAATGTGATCCATCTGCATATAGACAGCGTGTG 303  
DB 343 TGAGAGAAAAGTGACATGAACGACAGCTCTGCACTCGCCCTTTCAGGTGGCGCC 402  
QY 304 AAGAAGCTTCTCTTCTTGATCATCTGATTAACCTTAATTTGCTCTCAGCAAT 363  
DB 403 CGAAGGTGGTCTCTGCTGCTTCCATCTGATCTTCCCTGCAAGACAGCTACCCGATA 462  
QY 364 CTGAGTCTGCTTGTTAACAGCATCGCTTGAACCAACTAATGAGTGTATGCTATTGCT 423  
DB 463 GATGAGACCATGATCCAAATGGGCTCCCAACACAGCAATTTGGGTACTCGTATGCC 522  
QY 424 AGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGAGATGGAGTCAATC 483  
DB 523 AAGAGGATGATCGACGTGAGAAACAGGGCTACTTCCAGCAGTACGGCTGCACCTTCACC 582  
QY 484 TCTGGCATCGCTACTTAATCTCTATGTCCTTAATGACAAATTCACCCGAGATTCAT 543  
DB 583 GCTGTATCCCCCAACAGCTCTTCGGGGCCCCAGCAACTTCAACATCGAGATGCCAC 642

QY 544 GTGCTTCTGCTCTTTATGAGGAGGTTCCACGAGGGGAAAGTGAATGAGCGGAGAGTT 603  
DB 643 GTGCTGCTTGGCCTCATCCACAAGGTGCACCTTGGCCACAGCAGCGGCTCGGCC 699  
QY 604 GTGGTGTGGGTACAGGTAGTCCGTTGAGGAGTCTTTCATGTTGATGATTTGGCTGAT 663  
DB 700 ACGGTGTGGGTACAGGAATCCGCGGAGGAGTTCATATATCTCGTGTGACCTGGCCAG 759  
QY 664 GCTTGTGTTTCTTGTGATCGATACACAGCGGTTGGAGCATGTTA-----ACATTTGA 717  
DB 760 CTCCTTATCTGCTCTGCGGAGTACAAATGAAGTGGAGCCCATCATCTCTCCGTGGGC 819  
QY 718 AGTGTCAAGAGTCACTATTAGAGAGTTGGTGTGATGTTGGTGAAGAGGTTGTTGTTT 777  
DB 820 GAGGAAGATGAGTCTCCATCAAGGAGGAGCGGCGGTGTGGAGGCCATGAGCTTC 879  
QY 778 GAAGGGAAGCTTGGATGGATTGCACTAAGCCAGATGG 815  
DB 880 CATGGGAAGTCACTTTGATACAACCAAGTCGGATGG 917.

RESULT 7  
US-09-221-017B-669/c  
Sequence 669, Application US/09221017B  
Patent No. 644799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H.  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 669:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)

```
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...517
US-09-221-017B-669

Query Match      5.8%; Score 54.6; DB 4; Length 517;
Best Local Similarity 52.9%; Pred. No. 1.5e-08;
Matches 117; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 703 CATGTTAAACATTGGAAGTGGTCAAGAAGTCACTATTAGAGAGTTGGCTGAGTTGGTGAA 762
   |||||
Db 425 CATATCAACATTGGCACCAGGAGGAATCTCATTCGGATTGGCTTCGCTGATAGCA 366

QY 763 GAGGTTGTTGGTTTGAAGGAAGCTTTGGATGGATGCACTAAGCCAGATGGCACACG 822
   |||||
Db 365 AAGACCATCGTTTACGAGGCGCTTATCGAGTTTCGACTCTCTCAAAACCGACGCGATG 306

QY 823 AGGAACCTTATGACAGCTCAAGCTCGCTCTTTGGTTGGACACCTAAGCTTCTCTT 882
   |||||
Db 305 CGCAACTGACGATGTAAGCAAGCTCCATGCTTAGGCTTGAAGCATCGGATCGACATA 246

QY 883 AGAGATGGTCTGAGCCAAACTTATGATTGTTGTTGAAGAA 923
   |||||
Db 245 ACCACAGGTGACGAATGTTAGTGATTCGTCCGAA 205
```

## RESULT 8

```
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22113-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-P1s
US-08-232-463-14

Query Match      4.4%; Score 41.4; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 0.0026;
Matches 9; Conservative 214; Mismatches 160; Indels 0; Gaps 0;

QY 552 TGCTTTATGAGGAGTTCCAGGAGCGAAAGTGAATGAGCGGAGGAGATTGCTGGTGTG 611
   |||
Db 1440 TGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1381

QY 612 GGGTACAGTAGTCGGTTGAGGAGTCTTTCATGCTTGCATGTTGGCTGATGCTTGCT 671
   |||
Db 1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321

QY 672 TTTCTTGGTGCATACAGCGGTTGGAGCATGTTAAACATTGGAAGTGGTCAAGAAAT 731
   |||
Db 1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261

QY 732 GACTATTAGAGATTGGCTGAGTTGGTGAAGAGGTTGTTGTTTGAAGGAACTTGG 791
   |||
Db 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201

QY 792 ATGGATTGCACTAAGCCAGATGGCACACGAGAAACTTATGGACACTCAAAAGCTCGC 851
   |||
Db 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141

QY 852 GTCTTTGGTTGGACACCTAAGTTCTTCTTAGAGATGCTGCTGAGCCAAACTTATGATTG 911
   |||
Db 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081

QY 912 GTATTGGAAGTATTGTTCAACC 934
   |||
Db 1080 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1058
```

## RESULT 9

```
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22113-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
```





/ FILING DATE: 16-MAY-1979  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER:  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Dianne Casuto  
/ REGISTRATION NUMBER: P-40,943  
/ REFERENCE/DOCKET NUMBER: 4952.US.P2  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (847)-938-3137  
/ TELEFAX: (847)-938-2623  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 1:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 925 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: double  
/ TOPOLOGY: linear  
/ US-09-078-166-1

Query Match 3.9%; Score 36.6; DB 3; Length 925;  
Best Local Similarity 47.8%; Pred. No. 0.029;  
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 573 CGAGGCGAAAGTGAATGAGCGGAGGAAGTGTGGTGTGGGTACAGGTAGTCCGTTGAG 632  
DB 754 CGAGGCGGTGAGGTTGTGTCGGGCCGATTCGAGGTAGTCCGTTGAGTGTGT 695

QY 633 CGAGTTCTTGCATGTTGATGATTTGGCTGATGTTCTTCTTCTGCTGGATCGATACAG 692  
DB 694 CGAGTGTGAGGTTGCGGTGTAGCGGCGGTGCGGTGATGCTGGTCCAGTGGG 635

QY 693 CGGTTGGAGCATGTTAACTTGAAGTGGTCAAGAGTGCATTTAGAGAGTTGGCTGA 752  
DB 634 TGGTGTGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 575

QY 753 GTTGGTGAAGAGTTGTTGTTTGAAGGGAAGCTTGGATGGATT 799  
DB 574 GGTGGGTGAGGTTGTGAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 528

RESULT 12  
US-08-997-467-1/c  
/ Sequence 1, Application US/08997467  
/ Patent No. 6200813  
/ GENERAL INFORMATION:  
/ APPLICANT: Katz, Leonard  
/ APPLICANT: Stassi, Diane L.  
/ APPLICANT: Summers Jr., Richard G.  
/ APPLICANT: Ruan, Xiaon  
/ APPLICANT: Pereda-Lopez, Ana  
/ APPLICANT: Kakavas, Stephan J.  
/ TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES  
/ NUMBER OF SEQUENCES: 34  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Abbott Laboratories  
/ STREET: 100 Abbott Park Rd.  
/ CITY: Abbott Park  
/ STATE: Illinois  
/ COUNTRY: USA  
/ ZIP: 60064-3500  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: DOS  
/ SOFTWARE: FastSeq Version 2.0  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/997,467  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: 08/858,003  
/ FILING DATE: 16-MAY-1997  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Dianne Casuto  
/ REGISTRATION NUMBER: P-40,943  
/ REFERENCE/DOCKET NUMBER: 4952.US.P2  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (847)-938-3137  
/ TELEFAX: (847)-938-2623  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 1:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 925 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: double  
/ TOPOLOGY: linear  
/ US-08-997-467-1

Query Match 3.9%; Score 36.6; DB 4; Length 925;  
Best Local Similarity 47.6%; Pred. No. 0.029;  
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 573 CGAGGCGAAAGTGAATGAGCGGAGGAAGTGTGGTGTGGGTACAGGTAGTCCGTTGAG 632  
DB 754 CGAGGCGGTGAGGTTGTGTCGGGCCGATTCGAGGTAGTCCGTTGAGTGTGT 695

QY 633 CGAGTTCTTGCATGTTGATGATTTGGCTGATGTTCTTCTTCTGCTGGATCGATACAG 692  
DB 694 CGAGTGTGAGGTTGCGGTGTAGCGGCGGTGCGGTGATGCTGGTCCAGTGGG 635

QY 693 CGGTTGGAGCATGTTAACTTGAAGTGGTCAAGAGTGCATTTAGAGAGTTGGCTGA 752  
DB 634 TGGTGTGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 575

QY 753 GTTGGTGAAGAGTTGTTGTTTGAAGGGAAGCTTGGATGGATT 799  
DB 574 GGTGGGTGAGGTTGTGAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 528

RESULT 13  
US-08-149-476-281  
/ Sequence 281, Application US/09149476  
/ Patent No. 6420526  
/ GENERAL INFORMATION:  
/ APPLICANT: Rosen et al.  
/ TITLE OF INVENTION: 186 Human Secreted proteins  
/ FILE REFERENCE: PZ002P1  
/ CURRENT APPLICATION NUMBER: US/09/149,476  
/ EARLIER FILING DATE: 1998-09-08  
/ EARLIER APPLICATION NUMBER: PCT/US98/04493  
/ EARLIER FILING DATE: 1998-03-06  
/ EARLIER APPLICATION NUMBER: 60/040,162  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/040,333  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/038,621  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/040,626  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/040,334  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/040,336  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/040,163  
/ EARLIER FILING DATE: 1997-03-07  
/ EARLIER APPLICATION NUMBER: 60/047,600  
/ EARLIER FILING DATE: 1997-05-23  
/ EARLIER APPLICATION NUMBER: 60/047,615  
/ EARLIER FILING DATE: 1997-05-23  
/ EARLIER APPLICATION NUMBER: 60/047,597  
/ EARLIER FILING DATE: 1997-05-23  
/ EARLIER APPLICATION NUMBER: 60/047,502  
/ EARLIER FILING DATE: 1997-05-23





EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 3.7%; Score 34.8; DB 4; Length 1990;

Best Local Similarity 48.8%; Pred. No. 0.2;

Matches 124; Conservative 0; Mismatches 127; Indels 3; Gaps 1;

QY 592 GCGAGGAAGTTGTGGTGGGTACAGGTAGTCCGTTCCAGGGAGTTCTTGATGTTGAT 651  
DB 871 GGGAGCCACTCAGGTATACGATCCGGTCTCAGACAGGCGTTCAGTACGTGACG 930  
QY 652 GATTTGGCTGATCTTGTGTTTCTTGTGGATCGATACAGGGGTTGGAGCATGTTAAC 711  
DB 931 GATCTAGTGAATGGCTCGTGGCTCTCATGAACAG--CAAGTCAGCAGCCCGTCAAC 987  
QY 712 ATTGGAAGTGGTCAAGACTACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTT 771  
DB 988 CTGGGAACCCAGAGAACACACAAATCTTAGAATTTGCTCAGTTAATTAAGAAACCTGTT 1047  
QY 772 GGTTTTGAAGGAAGCTTGGATGGGATTCACATTAAGCCAGATGGCACACCGAGGAACCTT 831  
DB 1048 GTAGCGGAAGTGAATTCAGTTCTCTCGAAGCCAGGATGACCCACAGAAAGAAA 1107  
QY 832 ATGCACAGCTCAAA 845  
DB 1108 CCAGACATCAAAA 1121

## RESULT 14

US-09-149-476-131  
Sequence 131, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002P1  
CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163

EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,584  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,500  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,587  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,492  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,598  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,613  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,582  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,596  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,674  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,669  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,312  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056,886  
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,877  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670

EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 3.7%; Score 34.8; DB 4; Length 2041;  
Best Local Similarity 48.8%; Pred. No. 0.2;  
Matches 124; Conservative 0; Mismatches 127; Indels 3; Gaps 1;

QY 592 GCGGAGGAAGTTGTGTGGGTACAGGTAGTCCGTTGAGGAGTCTTTCATGTTGAT 651  
DB 894 GGGGAGCCACTCAGGTATACGGATCCGGTCTCAGACAAGGGCGTTCCAGTACGTACG 953  
QY 652 GATTTGGCTGATCTGTGTTTCTTCTGCTGGATCGATACAGCGGGTTGGAGCATGTAAC 711  
DB 954 GATCTAGTGAATGGCTCTGCTGCTCTCATGAACAG---CAACGTACAGCCCGGTCAAC 1010  
QY 712 ATTGGAAGTGGTCAAGAAGTGAATTTAGAGAGTTGGCTCAGTTGTTGAAAGAGTTGTT 771  
DB 1011 CTGGGGAACCCAGAGAACAACATCTTAGAATTTGCTCAGTTAATTAACCTTTGTT 1070  
QY 772 GGTTTTGAAGGAAGCTTGGATGGGATTCACCTAAGCCAGATGGCACACCGAGAACTT 831  
DB 1071 GGTAGCGGAAGTGAATTCAGTTTCTCTCGAAGCCAGGATGACCCACAGAAAAGAAA 1130  
QY 832 ATGGACAGCTCAAA 845  
DB 1131 CCAGACATCAAAA 1144

RESULT 15  
US-09-738-894A-3/c  
Sequence 3, Application US/09738894A  
Patent No. 6331423  
GENERAL INFORMATION:  
APPLICANT: GUEGLER, Karl et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEAR  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CL000636  
CURRENT APPLICATION NUMBER: US/09/738,894A  
CURRENT FILING DATE: 2000-12-18  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 36651

```
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(36651)
; OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match      3.4%; Score 32.2; DB 4; Length 36651;
.Best Local Similarity 48.6%; Pred. No. 10;
Matches 88; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 498 TAATCTCTATGGTCCTAATGACAAATTCACCCGAGAAATCTCATGTGCTTCCTGCTCT 557
Db 35539 TTATATTGTATCCCAATATTTTTCGCAGCTGTAGAATTCACCTTGGCTGTGCTTT 35480

Qy 558 TATGAGGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGAAAGTTGTGTGTGGGTAC 617
Db 35479 TGAGAAAGTGCTAATTTCTACCATACAGTGTGGCTGGCGAGCCACTTGAATGATGGTTAG 35420

Qy 618 AGGTAGTCCGTTGAGGGAGTCTTTCATGTTGATGATTGGCTGATGCTTTGTTTCTT 677
Db 35419 TATTTTTCATTCAAATTCCTTCTGTTTTTTTGTGTTTGTGTTTGTGTTTGTGTTT 35360

Qy 678 G 678
Db 35359 G 35359
```

Search completed: June 3, 2003, 04:45:32  
Job time : 96 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2003, 02:35:46 ; Search time 291 Seconds  
(without alignments)  
7243.541 Million cell updates/sec

Title: US-10-089-014-2

Perfect score: 936

Sequence: 1 atgtctgcaaatctgcacaa.....tgagaagtgttgcaaccca 936

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	936	100.0	936	22	AAH45734
2	857	91.6	1583	21	AAZ45315
3	659	70.4	1353	21	AAZ45315
4	391	41.8	1490	22	AAH88781
5	297	31.7	1020	18	AAH88781
6	297	31.7	4435	18	AAH88781
7	295.4	31.6	1020	18	AAH88781
8	295.4	31.6	7995	18	AAH88781
9	276.8	29.6	4403765	22	AAI99683
					Arabidopsis thaliana
					DNA encoding a GDP
					Arabisopsis thaliana
					GDP-4-keto-6-deoxy
					Open reading frame
					Part of the GS reg
					Open reading frame
					Complete GS region
					Mycobacterium tube

10	275.2	29.4	12732	21	ABQ62492	Mycobacterium tube
11	275.2	29.4	4411529	22	AAI99682	Mycobacterium tube
12	269.8	28.8	14024	20	AAZ06749	E. coli O157 anti-g
13	269.8	28.8	14024	21	AAZ56386	Escherichia coli f
14	265	28.3	534720	19	AAV30458	Rhizobium species
15	265	28.3	536165	19	AAV30459	Rhizobium species
16	262.2	28.0	966	21	AAZ45316	DNA encoding a GDP
17	258.8	27.6	966	21	AAZ45316	M. capsulatus gene
18	228.6	24.4	2182	23	AAH82555	DNA encoding novel
19	213.6	22.8	701	22	AAH88715	GDP-4-keto-6-deoxy
20	181.6	19.4	735	23	AAH82551	DNA encoding novel
21	179.4	19.2	14516	21	AAZ56375	E. coli O111 anti-g
22	179.4	19.2	14516	21	AAZ56375	Escherichia coli f
23	157.8	16.9	783	23	AAH82552	DNA encoding novel
24	157.8	16.9	2160	23	AAH82552	DNA encoding novel
25	145.8	15.6	933	24	AAI43590	Helicobacter pylori
26	139.2	14.9	2082	23	AAH89647	Helicobacter pylori
27	139.2	14.9	2410	23	AAH89647	DNA encoding novel
28	123.2	13.2	572	24	ABQ19306	Oligonucleotide fo
29	123.2	13.2	572	24	ABQ19307	Oligonucleotide fo
30	114	12.2	642	18	AAH68057	H. pylori cytoplas
31	104.8	11.2	1316	24	ABK70075	Antibody productio
32	103.2	11.0	1352	20	AAH89693	Nucleic acid seque
33	101.6	10.9	1340	21	AAZ45317	DNA encoding a GDP
34	93.4	10.0	3498	23	AAH64320	DNA encoding novel
35	86.4	9.2	572	24	ABQ19308	Oligonucleotide fo
36	86.4	9.2	572	24	ABQ19309	Oligonucleotide fo
37	78.6	8.4	999	23	ABL06261	Drosophila melanog
38	67.2	7.2	1638	23	AAH77377	DNA encoding novel
39	67.2	7.2	1638	23	AAH77377	DNA encoding novel
40	59.8	6.4	282	24	ABL72684	Corn tassel-deriva
41	58.8	6.3	447	21	AAA78009	CDNA encoding huma
42	58.8	6.3	447	21	AAI28747	Colon tumour relat
43	57.6	6.2	234	18	AAH75556	H. pylori cytoplas
44	53	5.7	2891	23	ABL22906	Drosophila melanog
45	53	5.7	3167	23	ABL06260	Drosophila melanog

#### ALIGNMENTS

RESULT 1  
AAH45734  
ID AAH45734 standard; DNA; 936 BP.

XX AC AAH45734;

XX DT 06-SEP-2001 (first entry)

XX DE A thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase gene.

XX KW GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase; GDP-L-fucose;

XX KW sugar chain; ds.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers

FT CDS 1..936

FT /tag= a

FT /product= "GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-

FT 4-reductase"

FT /partial

XX PN WO200138507-A1.

XX PD 31-MAY-2001.

XX XX 30-MAR-2000; 2000WO-JP02049.

XX PR 19-NOV-1999; . 99JP-0329045.

XX XX (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.

XX

PI Nakayama K, Jigami Y;  
 XX WPI; 2001-381292/40.  
 DR P-PSDB; AAG62614.  
 XX  
 CC Recombinant GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase  
 PT derived from Arabidopsis for efficient production of GDP (guanine  
 PT diphosphate)-L-fucose in vitro or in vitro  
 XX  
 PS Claim 3; Page 27-29; 35pp; Japanese.  
 XX  
 CC The present invention provides the protein and coding sequences of the  
 CC Arabidopsis thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-  
 CC reductase. The sequence can be used for the efficient production of  
 CC GDP-L-fucose, which is essential to the production of functionally  
 CC important sugar chains containing L-fucose. The present sequence is the  
 CC coding sequence of the invention.  
 XX  
 SQ Sequence 936 BP; 227 A; 190 C; 240 G; 279 T; 0 other;  
 Query Match 100.0%; Score 936; DB 22; Length 936;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-291;  
 Matches 936; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGCTGACAAATCTGCCAAATCTTGTGCGGGTCACTGCTGTTGGTGGTATCTGCC 60  
 DB 1 ATGCTGACAAATCTGCCAAATCTTGTGCGGGTCACTGCTGTTGGTGGTATCTGCC 60  
 QY 61 ATTGTCGCGAAGCTTCAGGAACAAGTTTCCAAATCTGTTCTTTAAACACACGCGGAG 120  
 DB 61 ATTGTCGCGAAGCTTCAGGAACAAGTTTCCAAATCTGTTCTTTAAACACACGCGGAG 120  
 QY 121 CTTGATCTCACTCGTCAAGCGGATTTGAATCTCTTTTCTCAAGAGAGCCAGTTTAT 180  
 DB 121 CTTGATCTCACTCGTCAAGCGGATTTGAATCTCTTTTCTCAAGAGAGCCAGTTTAT 180  
 QY 181 GTAATCCTAGCAGCAGTAAAGTTGGTGGTATTACGCTAACACACCTATCTCTGCTGAT 240  
 DB 181 GTAATCCTAGCAGCAGTAAAGTTGGTGGTATTACGCTAACACACCTATCTCTGCTGAT 240  
 QY 241 TTCAATGGTGTCAATCTCCAGATTCAGCAATGTGATCCACTCTGCATATGAGCAGGT 300  
 DB 241 TTCAATGGTGTCAATCTCCAGATTCAGCAATGTGATCCACTCTGCATATGAGCAGGT 300  
 QY 301 GTGAAGAAGCTTCTCTTCTTGGATCATCTGATTTACCTAAATTTGCTCTCAGCCA 360  
 DB 301 GTGAAGAAGCTTCTCTTCTTGGATCATCTGATTTACCTAAATTTGCTCTCAGCCA 360  
 QY 361 ATTCCTGAGTCTCTTGTGTTTAAACAGCATCGCTTGAACCACTAATGAGTGGTATGCTATT 420  
 DB 361 ATTCCTGAGTCTCTTGTGTTTAAACAGCATCGCTTGAACCACTAATGAGTGGTATGCTATT 420  
 QY 421 GCTAAGATCGCTGGGATTAAGCTTGTGAGGCTTATAGGATTCAGCAGCGATGGGATGCA 480  
 DB 421 GCTAAGATCGCTGGGATTAAGCTTGTGAGGCTTATAGGATTCAGCAGCGATGGGATGCA 480  
 QY 481 ATCTCTGGCATGCTACTAATCTCTATGTCCTAATGACAATTTCCACCGGAGATTCT 540  
 DB 481 ATCTCTGGCATGCTACTAATCTCTATGTCCTAATGACAATTTCCACCGGAGATTCT 540  
 QY 541 CATGCTCTTCTCTCTTATGAGGAGTTCCACGAGCGGAAAGTGAATGAGCGGAGGAA 600  
 DB 541 CATGCTCTTCTCTCTTATGAGGAGTTCCACGAGCGGAAAGTGAATGAGCGGAGGAA 600  
 QY 601 GTTGTGGTGGGATCAGGTAGTCTGTTGAGGAGTTCTTGCATGTTGATTTGGCT 660  
 DB 601 GTTGTGGTGGGATCAGGTAGTCTGTTGAGGAGTTCTTGCATGTTGATTTGGCT 660  
 QY 661 GATGCTTGTGTTTCTTCTGTTGATCGATACAGCGGTTTGAGCATGTTTAAACATTTGAAGT 720  
 DB 661 GATGCTTGTGTTTCTTCTGTTGATCGATACAGCGGTTTGAGCATGTTTAAACATTTGAAGT 720  
 QY 721 GGTCAAGAAGTGAATTAAGAGATTGGTGGTGAATGAGGTTTGTGTTTGA 780

721 GGTCAAGAAGTGAATTAAGAGATTGGTGGTGAATGAGGTTTGTGTTTGA 780  
 781 GGTCAAGAAGTGAATTAAGAGATTGGTGGTGAATGAGGTTTGTGTTTGA 840  
 841 TCAAGAGTCTGCTCTTGGTGGTGGACACCTAAGGTTTCTTTAGAGATGCTCTGAGCCAA 900  
 901 ACTTATGATGTTGTTTGAAGATGTTTGAACCGA 936  
 901 ACTTATGATGTTTGAAGATGTTTGAACCGA 936

RESULT 2  
 AA245315  
 ID AA245315 standard; DNA; 1583 BP.  
 XX  
 AC AA245315;  
 XX DT 27-MAR-2000 (first entry)  
 XX  
 DE DNA encoding a GDP-4-keto-6-deoxy-D-mannose epimerase/reductase.  
 XX  
 KW GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;  
 KW GDP-L-galactose; vitamin C; ascorbic acid; L-ascorbic acid;  
 KW ascorbic acid pathway enzyme; hexokinase; glucose phosphate isomerase;  
 KW phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;  
 KW GDP-D-mannose pyrophosphorylase; GDP-D-mannose-GDP-L-galactose epimerase;  
 KW GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;  
 KW L-galactono-gamma-lactone dehydrogenase; ester; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 49..993  
 FT /\*tag= a  
 FT /product= "GDP-4-keto-6-deoxy-D-mannose epimerase/  
 FT reductase"  
 XX  
 PN WO9964618-A1.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 26-MAY-1999; 99WO-US11576.  
 XX  
 PR 08-JUN-1998; 98US-0088549.  
 PR 17-MAR-1999; 99US-0125073.  
 PR 18-MAR-1999; 99US-0125054.  
 XX  
 PA (DCVB-) DCV INC DBA BIO-TECH RESOURCES.  
 XX  
 PI Berry A, Running JA, Severson DK, Burlingame RP;  
 XX  
 DR WPI; 2000-105890/09.  
 DR P-PSDB; AAY5414.  
 XX  
 PT Production of ascorbic acid or esters, using microorganisms or plants  
 PT which have genetic modification in enzymes involved in the ascorbic  
 PT acid synthesis pathway  
 XX  
 PS Claim 26; Page 164-166; 187pp; English.  
 XX  
 CC The present sequence encodes a GDP-4-keto-6-deoxy-D-mannose epimerase/  
 CC reductase. The enzyme catalyses the conversion of GDP-D-mannose to  
 CC GDP-L-galactose. The enzyme can be modified, and used to produce  
 CC transgenic microorganisms, which can be used in fermentation techniques  
 CC to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is  
 CC modified to increase its action. Other ascorbic acid pathway enzymes  
 CC which may be used in the method of the invention include hexokinases,  
 CC glucose phosphate isomerases, phosphomannose isomerases,

CC phosphomannomutases, GDP-D-mannose pyrophosphorylases,  
CC GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,  
CC L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and  
CC L-galactono-gamma-lactone dehydrogenases. The methods can be used for  
CC the production of ascorbic acid or esters using microorganisms or plants.  
XX  
SQ Sequence 1583 BP; 407 A; 330 C; 351 G; 495 T; 0 other;

Query Match 91.6%; Score 857; DB 21; Length 1583;  
Best Local Similarity 97.1%; Pred. NO. 1.3e-265;  
Matches 918; Conservative 0; Mismatches 15; Indels 12; Gaps 4;  
QY 1 ATGCTGACAAATCTGCCAAATCTTCGTGGGGTCATCGTGGTTGGTTGGATCTGCC 60  
DB 49 ATGCTGACAAATCTGCCAAATCTTCGTGGGGTCATCGTGGTTGGTTGGATCTGCC 108  
QY 61 ATTGTCGGCAAGCTTCAGGAACAAGGTTTCACCAATCTCGTTCTTAAACACACGCCGAG 120  
DB 109 ATTGTCGGCAAGCTTCAGGAACAAGGTTTCACCAATCTCGTTCTTAAACACACGCCGAG 168  
QY 121 CTTGATCTCACTCGTCAAGCCGATGTTGAATCTTCTTCTCAAGAGAAGCCAGTTTAT 180  
DB 169 CTTGATCTCACTCGTCAAGCCGATGTTGAATCTTCTTCTCAAGAGAAGCCAGTTTAT 228  
QY 181 GTAACTCTAGCAGCAGCTAAAGTTGGTGTATTCAGCTAAACAACCTATCTCTGCTGAT 240  
DB 229 GTAACTCTAGCAGCAGCTAAAGTTGGTGTATTCAGCTAAACAACCTATCTCTGCTGAT 288  
QY 241 TTCAATGGTGTCAATCTCAGATTCAGACCAATGTGATCCACTTCGCATATGAGCAGGT 300  
DB 289 TTCAATGGTGTCAATCTCAGATTCAGACCAATGTGATCCACTTCGCATATGAGCAGGT 348  
QY 301 GTGAAGAAGCTTCTTCTCTGATCATCTGCTGATCACTTACCTTAAATTTCTCTCAGCCA 360  
DB 349 GTGAAGAAGCTTCTTCTCTGATCATCTGCTGATCACTTACCTTAAATTTCTCTCAGCCA 408  
QY 361 ATTCCTGAGTCTGCTTGTATTAAGCAGATCGTGTGAACCACTTAATGAGTGTATGTTAT 420  
DB 409 ATTCCTGAGTCTGCTTGTATTAAGCAGATCGTGTGAACCACTTAATGAGTGTATGTTAT 468  
QY 421 GCTAAGATCGTGGGATTAAGACTTCTGAGGCTTATAGGATTCAGCAGCGATGGATGA 480  
DB 469 GCTAAGATCGTGGGATTAAGACTTCTGAGGCTTATAGGATTCAGCAGCGATGGATGA 528  
QY 481 ATCTCTGGCATGCTACTAATCTCTATGGTCTTAATGACAAATTTCCACCGGAGATTTCT 540  
DB 529 ATCTCTGGCATGCTACTAATCTCTATGGTCTTAATGACAAATTTCCACCGGAGATTTCT 585  
QY 541 CATGCTCTTCTGCTTATAGGAGGTTCCAGGAGCGAAAGTGAA-TGGAGCGGAGGA 599  
DB 586 CATGCTCTTCTGCTTATAGGAGGTTCCAGGAGCGAAAGTGAAATGGAGCGGAGGA 645  
QY 600 AGTTGTGGTGGGGTACAGGTAGTCC- - - - -GTTGAGGGAGTTCCTGATGTTGATGAT 654  
DB 646 AGTTGTGGTGGGGTACAGGTAGTCCCGTTGGAAGGGAGTTCCTGATGTTGATGAT 705  
QY 655 TTGGCTGATGCTTGTGTTTCTTCTGCTGATCG- - - - -ATACAGCGGTTGGAGCATGTTAAC 711  
DB 706 TTGGCTGATGCTTGTGTTTCTTCTGCTGATCGATACAGCGGGTTGGAGCATGTTAAC 765  
QY 712 ATTGGAAGTGGTCAAGAAGTGAATATAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTT 771  
DB 766 ATTGGAAGTGGTCAAGAAGTGAATATAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTT 825  
QY 772 GTTTTTGAAGGAAGCTTCGATGGATTCACCTAAGCCAGATCGCACCGGGAACCTT 831  
DB 826 GTTTTTGAAGGAAGCTTCGATGGATTCACCTAAGCCAGATCGCACCGGGAACCTT 885  
QY 832 ATGGACAGCTCAAAGCTTCGCTCTTGGGTGGGATGGACACCTTAAGGTTTCTCTTAGAGATGT 891  
DB 886 ATGGACAGCTCAAAGCTTCGCTCTTGGGTGGGATGGACACCTTAAGGTTTCTCTTAGAGATGT 945  
QY 892 CTGAGCCAAACTATGATGGTATTTGAAGAATGTTTGAACCGA 936

DB 946 CTGAGCCAAACTATGATGGTATTTGAAGAATGTTTGAACCGA 990  
RESULT 3  
AAC34827  
ID AAC34827 standard; DNA; 1353 BP.  
XX AAC34827;  
AC AAC34827;  
XX 17-OCT-2000 (first entry)  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 8033.  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.

```
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139819.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 659; DB 21; Length 1353;
Best Local Similarity 82.6%; Pred. No. 9.1e-202;
Matches 766; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCACTCGTGGTTTGGTGGATCTGCC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 199 ATGTTAGAGAAATCGCCAAATCTTTTCTGCGCAATAGGGGATTCGTTGGATCCGCC 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 ATTGTCCGCAAGCTTCAGGAACAAGTTTTCACCAATCTCGTTCTTTAAACACACGCCGAG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 ATAGTCCGCAAGCTTCAGGATCAGGGTTTTCACCAATCTCGTTCTTCGAAACATTCGAG 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



QY 121 CTTGATCTCAGTCCGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTAT 180  
Db 319 CTTGATCTCAGTCCGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTAT 378  
QY 181 GTAATCCTAGCAGCAGTAAAGTTGGTGTATTCAGCTAAACAACACCTATCTCTGAT 240  
Db 379 GTCATCTCTGCTGGGCTAAAGTTCGGTGAATTCATGCCAACACACATACCAGCTAT 438  
QY 241 TTCAATGGTGTCAATCTCAGATTCAGACCAATGTGATCCACTCTGATATGAGCAGCGT 300  
Db 439 TTCAATGGTGTCAATCTCAGATTCAGACCAATGTGATCCACTCTGATATGAGCAGCGT 498  
QY 301 GTGAGAACCTCTCTCTCTGATCATCTGATTAACCTTAATTTCTCTGATCA 360  
Db 499 GTGAGAACCTCTCTCTCTGATCATCTGATTAACCTTAATTTCTCTGATCA 558  
QY 361 ATTCTGATCTGCTTTGTTAAACAGCATCGCTTGAACCAATTAATGAGTGTATCTAT 420  
Db 559 ATTCCGATCTGCGCTTTACCGGTCGCTCGAACCGAACGAATGGTATGGATC 618  
QY 421 GCTAAGATCGCTGGGATTAAGACTTTGTAGGCTTATPAGGATTCAGCAGGATGGATCA 480  
Db 619 GCCAAGATCGCGGATTAAGATGTTCAAGCTTACCGGCTTCAGCATCAGTGGGATGCT 678  
QY 481 ATCTCTGGCATGCTACTTAATCTCTATGCTCTTAATGACAAATTCACCCGAGAAATCT 540  
Db 679 ATCTCTGGCATGCTACTTAATCTCTATGCTCTTAATGACAAATTCACCCGAGAAATCT 738  
QY 541 CATGTCTCTCTGCTCTTATGAGGAGTTTCCAGCGGCGAAAGTGAATGAGCGAGGAA 600  
Db 739 CATGTCTCTCTGCTCTTATGAGGAGTTTCAATGAGCTTAAAGCCAAATATCTGACGAA 798  
QY 601 GTTGTGTGTGGGTACAGTATGCTGTTGAGGAGTTCTTGATGTTGATTTGGCT 660  
Db 799 GTTGTGTGTGGGAGTGAAGCCGTTGAGGAGTTTACATGTGGATGTTGGCT 858  
QY 661 GATGCTGTGTTTCTTCTGATGATGATACAGCGGTTGGAGCATGTTAAACATTTGAAGT 720  
Db 859 GATGCTGTGTTTCTTCTGATGATGATACAGCGGTTGAGCATGTTGAGCATTTAGGAAT 918  
QY 721 GGTCAAGAGTCACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTCTGTTTGA 780  
Db 919 GGTGTTGAAGTCAAGATCAAGATTTGGCTGAGTTGGTGAAGAGGTTCTGTTTGA 978  
QY 781 GGGAGCTTTGGATGGATTTGACTAAGCCAGATGGCACACCAGGAGAACTTATGGACAGC 840  
Db 979 GGAAGCTTTGTTGGATACCAACAAACAGAGCGGACACCAAGAACTGATGATAC 1038  
QY 841 TCAAGCTCGCTCTTTGGGTTGGACACCTAAGTTTCTTCTAGAGATGGTCTGAGCCAA 900  
Db 1039 TCAAGCTTTGCTC-TTGGGATGGACACCGAAGATTTCTGCTAAAGATGGCTCTCCCA 1097  
QY 901 ACTTATGATTTGATTTGAAGATGTT 927  
Db 1098 ACTTATGATTTGATTTGAAGATGTT 1124

RESULT 4  
ID AAH88781 standard; cDNA; 1490 BP.  
XX  
AC AAH88781;  
XX  
DT 28-SEP-2001 (first entry)  
XX  
DE GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase-4-reductase cDNA sequence #74.  
XX  
KW Moss; carbohydrate metabolism related protein; CMRP; sugar; cofactor;  
KW fine chemical production; carbohydrate; polysaccharide; ss.  
XX  
OS Physcomitrella patens.  
XX  
PN WO200144476-A2.

XX 21-JUN-2001.  
XX 14-DEC-2000; 200WO-EP12697.  
XX 16-DEC-1999; 99US-0171101.  
XX (BADI ) BASF PLANT SCI GMBH.  
XX Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;  
XX Frank M, Freund A, Duwenig E, Schmidt R, Reeki R;  
XX WPI; 2001-398155/42.  
XX P-PSDB; AAM00095.  
XX Novel moss nucleic acid molecules encoding a carbohydrate metabolism  
XX as carbohydrates, cofactors and enzymes from microorganisms and plants  
XX  
XX Claim 7; Page 104; 133pp; English.  
XX This invention relates to nucleic acid molecules AAH88708 - AAH88796  
XX isolated from Physcomitrella patens (a moss), which encode carbohydrate  
XX metabolism related proteins (CMRP) represented in AAM00022 - AAM00110.  
XX Included in the invention is a vector containing the CMRP cDNA, and a  
XX host cell transformed with the vector. The host cell (a microorganism,  
XX Corynebacterium or Brevibacterium, moss or algae or a plant cell) is  
XX useful for producing a fine chemical such as carbohydrates, cofactors  
XX and/or enzymes. The nucleic acid molecules are suitable for modifying a  
XX carbohydrate production system in a host, e.g., microorganisms and  
XX plants. They are also useful to identify those DNA sequences and enzymes  
XX in other species which are useful to modify the biosynthesis of starch,  
XX cell wall polysaccharides and sugars. The nucleic acid molecules may be  
XX utilised in the genetic engineering of Corynebacterium glutanicum and the  
XX related Brevibacterium species and Acetobacter xylinum and Chlorella to  
XX make it a better or more efficient producer of one or more fine  
XX chemicals. Mutagenesis of one or more CMRPs may also result in CMRPs  
XX having altered activities which indirectly impact the production of one  
XX or more desired fine chemicals from plants. Primers AAH88705 - AAH80707  
XX are used in the sequencing of the CMRP cDNA sequences of the invention.  
XX  
SQ Sequence 1490 BP; 369 A; 308 C; 410 G; 402 T; 1 other;  
Query Match 41.8%; Score 391; DB 22; Length 1490;  
Best Local Similarity 65.0%; Pred. No. 2.9e-115;  
Matches 596; Conservative 0; Mismatches 315; Indels 6; Gaps 1;  
QY 7 GACAAATCTGCCAAATCTTCGTCGGGTCTATCGTGGTTTGGTGGATCTGCCATGTGC 66  
Db 356 GACAAGGACGCCAAGATCTTTGTTGTCGACACCGAGGTCTAGTAGTGGCTGTGTT 415  
QY 67 CGCAGCTTCAGGACACAGGTTTACCAATCTCGTCTTAAACACACCGCCGCTTGAT 126  
Db 416 CGTGTCTTGAAGAAGGATGGTTATACCAATTTGGTGTGAAGACTCATAAAGACTAGAT 475  
QY 127 CTCCTCTGTCAGCGCATGTTGAATCTCTTTCTCAAGAGAACGACGTTTATGTAATC 186  
Db 476 CTTACCCGTCAGCAGAGG-----AATTTTCGACACGGAACACCGCTACGTCATC 529  
QY 187 CTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAAACAACCTATCTCTGCTGATTCATT 246  
Db 530 CTAGCAGCTCGAAGTGGGAGCATTCAGCGAAACAGACTTACCTCTCAGAGTTCAAT 589  
QY 247 GGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGAGCACCGTGTGAAG 306  
Db 590 GCGGTGAATCTGCAGATCCAAACGAATGTCAATCGATGCTGTCTTACAAAGTCTCGGGTGAAG 649  
QY 307 AAGCTTCTCTCTCTTGGATCATCTTCCATTTTACCCTTAAATTTGCTCTCAGCAATTCCT 366  
Db 650 AAGCTTCTCTCTCTTGGCTCTCTCGTGTATCTACCAAGATTTGCCAGGTACCCATCGTT 709  
QY 367 GAGTCTGCTTTTGTAAACAGCATCGCTTGAACCAACTAATGATGATGATGCTATTGCTAAG 426



Db 704 GCGCGAATCTCGATGTCGAGCATGTCGGAGCGCATGCTGTCTCTTTTGGAACTT 763

Qy 689 ACAGCGGGTTGGAGCATGTTAACTTTGAAGTGGTCAAGAAGTGAATATTAGAGAGTTGG 748

Db 764 TCGATGGTCCGAACACGTCACAGTGGGACCGGGTCCGATCACAGCATTTAGCGAGATCG 823

Qy 749 CTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGAGCTTGGATGGGATTCACCTAAGC 808

Db 824 CAGACATGGTCTGCTACGGCGGTGGGCTACATCGCGGAAACACAGTTGGGATCCAACTAAAC 883

Qy 809 CAGATGGCACCGAGGAACCTTATGGACAGCTCAAGAGCTCGCTCTTTGGGTTGGACAC 868

Db 884 CCGATGGAAACCCGCGCAAACTATTGGACCTTCGCGCTACGCGAGTTGGGTTGGCGCC 943

Qy 869 CTAAGTTTCTCTTGAAGATGGTCTGAGCCAACTTATGATGGTATTTGAAGAATG 925

Db 944 CCGCAATCGCACTGAAGAGCGCATCGATGCAACGGTGTGCTGTACCGCACAAATG 1000

RESULT 6

ID AAT74479

XX AAT74479 standard; DNA; 4435 BP.

AC AAT74479;

XX

DT 11-MAR-1998 (first entry)

XX

DE Part of the GS region DNA sequence from M. paratuberculosis.

XX

KW GS; pathogenicity island; pathogenic protein; mycobacterial disease;

KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;

KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.

XX

OS Mycobacterium paratuberculosis.

XX

FH Key

FT Location/Qualifiers

FT 201..1232

FT /tag= a

FT /note= "encodes AAW21769"

FT 1172..2191

FT /tag= b

FT /note= "encodes AAW21771"

FT 2467..3189

FT /tag= c

FT /note= "encodes AAW21773"

FT 3335..4135

FT /tag= d

FT /note= "encodes AAW21775"

XX

PN W09723624-A2.

XX

PD 03-JUL-1997.

XX

PF 23-DEC-1996; 96WO-GB03221.

XX

PR 21-DEC-1995; 95GB-0026178.

XX

PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.

XX

PI Doran T, Ford J, Hermon-Taylor J, Loughlin M, Millar D;

PI Sumar N, Tizard M;

XX-

DR WPI; 1997-351061/32.

XX

PT New isolated pathogenicity island from mycobacteria - used to

PT develop products for detection, diagnosis, prevention and treatment

PT of mycobacteria infections

XX

PS Claim 5; Pages 43-44; 62pp; English.

XX

CC The present sequence represents a novel polynucleotide sequence

CC designated "GS". GS is a pathogenicity island of 8 kb of DNA comprising

CC a core region of 5.75 kb with multiple open reading frames (ORFs) and

CC an adjacent transmissible element of 2.5 kb. The ORFs, and also the

CC transmissible element, encode proteins which may be linked to

CC pathogenicity, such as providing receptors for cellular recognition. GS

CC was discovered and characterised using differential DNA analysis

CC technology. It is found within Mycobacterium paratuberculosis and it has

CC also been identified in Mycobacterium avium subsp. silvaticum. These

CC pathogenic mycobacteria cause chronic inflammation of the intestine and

CC Crohn's disease in humans. The protein products of the ORFs of GS can be

CC used for detecting mycobacteria or for diagnosing, treating or preventing

CC mycobacterial diseases. In particular they can be used as vaccines for

CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or

CC Johne's disease in animals.

XX

Qy Sequence 4435 BP; 904 A; 1289 C; 1288 G; 954 T; 0 other;

XX

Query Match 31.7%; Score 297; DB 18; Length 4435;

Best Local Similarity 58.2%; Pred. No. 1.1e-84;

Matches 522; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

Qy 29 TCGCGGTCATCGTGGTTGGTTCGATTCGCAATTCGCGAAGCTTCAGGAACAAGTT 88

Db 1275 TCGCGGTCATCGCGGGCTGGTGGCTCAGCGCTGATGATTTGAGGCGCGAGGGT 1334

Qy 89 TCACCAATCTCGTTCTTAAACACACGCGGAGCTTGATCTCACTCGTCAAGCGGATGTTG 148

Db 1335 TCACCAATCTCATTTGCGATCACGCGATGATGATCTGACGACCGGCGCAACGT 1394

Qy 149 AATCCTCTTTTCAAGAGAGCCAGTTTATGTAATCTAGCAGCAGCTAAAGTTGGTG 208

Db 1395 TTGATTTTGTGTCGAGCAACACACAGGTGATCATGATCGCGCGGCGCGCGGCG 1454

Qy 209 GTATTACGCTAACACACCTATCTCTGCTGATTTCATTTGGTGTCAATCTCCAGATTCAGA 268

Db 1455 GCATCATGGCGAATAACACCTATCCGCGGACTTCTTGTCGAAACCTCCGAATCCAGA 1514

Qy 269 CCAATGTGATCCACTCTGCATATGACACGGTGTGAAGAGCTTCTTCTCTTGGATCAT 328

Db 1515 CCAATTTGCTCGACGAGCTGTCGCGCTGTCGCGGCTCTCTTCTCGGTTCTG 1574

Qy 329 CCGTCAATTTACCTAAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCAT 388

Db 1575 CATGATCTACCCGAACTAGCTCCGCAACCTATCAAGAGTGTCTTTATGATGGCC 1634

Qy 389 CGCTTGAACCACTAATGAGTGTATGCTATTGTAAGATCGCTGGGATTAAGACTTGTG 448

Db 1635 CTTTGGAGCCCAACACGCGTATGCGATCGCAAGATCGCGGTATCTCTGCAAGTTC 1694

Qy 449 AGGCTTATAGGATTCAGCACGGATGGGATGCAATCTCTGGCATGCTTACTATCTCTATG 508

Db 1695 AGGCGGTTAGCGCGCAATATGGGCTGGGCTGATCTCTGCGATGCGGACTAACTCTACG 1754

Qy 509 GTCCTAATGACAAATTTCCACCGGAGAAATTCCTCATGTGCTTCTCTTATGAGGAGGT 568

Db 1755 GACCGCGGACAACTTCTCCCGTCCGGTCCGATCTTTGCGGGGCTCATCCGTCGAT 1814

Qy 569 TCCACGAGCGCAAAAGTGAATGGAGCGGAGAAATTTGTTGTTGGGTACAGGTAGTCCGT 628

Db 1815 ATGAGGAGCCAAAGCTGGTGGTGCAGAAAGTGAAGTGAAGTGGGGACCGGTACTCCG 1874

Qy 629 TGAGGAGTCTTGTGATGTTGCTGATGTTGGCTGATGCTTGTGTTTCTTCTGCGATCGAT 688

Db 1875 GCGCGCAACTTCTGATGTCGACGATCTGGCGAGCGCATGCTTCTCTTTTGAACAAT 1934

Qy 689 ACAGCGGGTTGGAGCATGTTTAACTTTGAAGTGGTCAAGAAGTGAATTTAGAGAGTTGG 748

Db 1935 TCGATGGTCCGAAACCAACGTCACGTCGCGGCGATCACAGCATTAGCAGATCG 1994

Qy 749 CTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGAGCTTGGATGGGATTCACCTAAGC 808

Db 1995 CAGACATGGTCTACGGCGGTGGGCTACATCGCGCAACAGTTGGGATCCAACTAAAC 2054

Qy 809 CAGATGGCACACCGAGGAACCTTATGGACAGCTCAAGAGCTCGGCTTTGGGTTGGACAC 868









CC using a bacterial artificial chromosome (BAC) vector. Recombinant BAC  
 CC vectors, which are preferably immobilised, can be used to detect  
 CC mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological  
 CC samples. The polynucleotides identified are useful as probes or primers  
 CC for detecting a given mycobacterium of interest. By aligning the  
 CC polynucleotides contained in the recombinant BAC vectors it is possible  
 CC to physically map a polynucleotide of mycobacterial origin in a  
 CC biological sample. The methods and vectors from the present invention  
 CC are useful in providing information for combating tuberculosis. It is  
 CC possible to compare genomes between different strains or species and  
 CC their non-pathogenic strains or species counterparts. ABO62492 to  
 CC ABO63228 and ABB81227 to ABB81230 represent sequences used in the  
 CC exemplification of the present invention.

XX  
 SQ Sequence 12732 BP; 2342 A; 3849 C; 4091 G; 2450 T; 0 other;

Query Match 29.4%; Score 275.2; DB 21; Length 12732;  
 Best Local Similarity 56.2%; Pred. No. 2.1e-77;  
 Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

QY 6 TGACAAATCTGCACAAATCTTCGTCGGGTATCGTGGTATCTGTTGGTGGATCTGCCATTGT 65  
 DB 8106 TGACCGCGCGCGCGGTCTACATCGCCGGATCGCGCTGCTGGTCCGGTCCGGCTGT 8165

QY 66 CGCAGCTTCAGAACAGGTTTACCAATCTCTGTTCTTAAACACACGCCGAGCTTGA 125  
 DB 8166 AGCAGCTTTCGGCGCGCGGGTTTACCACTCTGTTGCGGTGCGCGCGCGAGCTGA 8225

QY 126 TCTCACTCTCAAGCGATGTTGAATCTTCTTCTCAAGAGAACCGATTATGTAAT 185  
 DB 8226 TCTGACGATCGGCGCGACGTTTCGACTTCTGTCGATCGAGCCGCGAGTGTCAAT 8285

QY 186 CTTAGCAGCAGCTAAAGTTGGTGGTATCAAGCTAACACACACTATCTCTGCTGATTAT 245  
 DB 8286 CGACGCGCGCGCGCGGTTCGGGCGATCTTGGCCACAGCACCTACCGCGCGATTCTCT 8345

QY 246 TGGTGTCAATCTCCAGATTCAGACCAATGTGATCTCACTCTGCATATGACACGGTGA 305  
 DB 8346 GTCCGAAACCTCCAGATTCAGATCACTGCTGATGCGCGGTGCGCGCGCGGTGCG 8405

QY 306 GAAGCTTCTCTCTTGGATCATCTGCTGATTTACCTTAATTTGCTCTCAGCAATCTCC 365  
 DB 8406 GGGGTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8465

QY 366 TGAGTCTGCTTTTAAACAGCATCGCTTGAACCACTTAATGATGTTGTTGTTGCTTAA 425  
 DB 8466 GGAGAGCGCGCTGCTCAACCGTCCGTTGGAGCGCAACAGCGGTACGCGATCGCAA 8525

QY 426 GATCGTGGGATTAAGACTTGTGAGCTTATAGGATTCAGACGATGGATGCAATCTC 485  
 DB 8526 AATCGCGCGATCTTTCGCGTCCAGCGGTGCGCGCGCAACATGCGCTCGGATCTC 8585

QY 486 TGGCATGCTCTACTTAATCTGCTCTTAATGACATTTCCACCGGAGAAATCTCATCT 545  
 DB 8586 GGGGATGCCCACTGATCGGCGCAGCGGCACTTTTTCGCGTCCGGTCCGGTCTGCT 8645

QY 546 GCTTCTGCTCTTATGAGAGGTTTCAACGAGCGGAAAGTGAATGAGCGGAGAACTGT 605  
 DB 8646 GCTCGCGCACTCATCGCGCTTATGACGAGGCGCAAGCGAGTGGCGCGCAACGTCAG 8705

QY 606 GGTGCGGTACAGGTAGTCTGTTGAGGAGTCTTTCGATGTTGATGTTGCTGATCG 665  
 DB 8706 CAACTGGGCGACCGCGCGCGCGGAGTTGCTGCACTGCGACGCTCGCGAGCGC 8765

QY 666 TTGTGTTTCTTCTGCTGATTCATACAGCGGTGAGGATGTTAACTTGGAGTGGTCA 725  
 DB 8766 ATGCTGTATCTCTGGAACATTTGACCGCGCGCGACCCCATGTCAACGTGGAGACCGGAT 8825

QY 726 AGAAGTGAATTTAGAGAGTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGAA 785  
 DB 8826 CGACACACCATCGCGAGATCGCGGAGATGTTGCTGCGGTAGGTATAGCGCGA 8885

QY 786 GCTTGGATGGATTCACATGACGATGGCAACCGAGGAAACTTTATGACAGCTCAAA 845

DB 8886 ABCCCGCTGGGATCCAAAGCAAAACCGGACGGAACACACCAACTCTGATGTTTCGGT 8945  
 QY 846 GCTCGCGCTTTTGGGTTGGACACCTAAAGGTTTCTCTTAGAGATGTTCTGAGCCAAACTTA 905  
 DB 8946 GCTACGGGAGCGGATGCGCGCTTCGATCGGCTGCGGCGGATCGAGCGACGCT 9005

QY 906 TGAATGGTATTTGAAGATG 925  
 DB 9006 GCGGTGATTCGAGCAGC 9025

RESULT 11  
 AAI99682  
 ID AAI99682 standard; DNA; 4411529 BP.  
 XX  
 AC AAI99682;  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
 XX  
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; Genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN US6294328-B1.  
 XX  
 PD 25-SEP-2001.  
 XX  
 PF 24-JUN-1998; 98US-0103840.  
 XX  
 PR 24-JUN-1998; 98US-0103840.  
 XX  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
 XX WPI; 2001-647261/74.  
 DR  
 XX  
 XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ  
 XX  
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
 XX  
 CC The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.  
 XX  
 SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Query Match 29.4%; Score 275.2; DB 22; Length 4411529;  
 Best Local Similarity 56.2%; Pred. No. 4.7e-76;  
 Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

QY 6 TGACAAATCTGCACAAATCTTCGTCGGGTATCGTGGTATCTGTTGGTGGATCTGCCATTGT 65  
 DB 1704120 TGACCGCGCGCGCGGTCTACATCGCGGCGATCGCGGCTGTTGGTCCGGCTGCT 1704179

QY 66 CCGCAAGCTTCAGGAACAAGGTTTACCAATCTCTTAAACACACACCGCGAGCTTGA 125









```
FT /note= "homologous to the Tnp gene"  
FT complement (433880..434110)  
FT *tag= k  
FT /standard_name= "ORF K11"  
FT /product= "protein of unknown function"  
FT /note= "homologous to the FixU gene"  
FT complement (434107..434433)  
FT *tag= l  
FT /standard_name= "ORF K12"  
FT /product= "protein of unknown function"  
FT complement (434517..434711)  
FT *tag= m  
FT /standard_name= "ORF K13"  
FT /product= "ferrodoxin/ferrodoxin-like protein"  
FT /note= "homologous to the FdxN gene"  
FT complement (434753..436234)  
FT *tag= n  
FT /standard_name= "ORF K14"  
FT /gene= "nifB"  
FT /product= "protein involved in FeMo co-factor  
FT biosynthesis"  
FT complement (436460..438130)  
FT *tag= o  
FT /standard_name= "ORF K15"  
FT /gene= "nifA"  
FT /product= "positive regulator of nif, fix and other  
FT genes"  
FT complement (438297..438590)  
FT *tag= p  
FT /standard_name= "ORF K16"  
FT /gene= "fixX"  
FT /product= "protein required for nitrogenase activity"  
FT complement (438605..439912)  
FT *tag= q  
FT /standard_name= "ORF K17"  
FT /gene= "fixC"  
FT /product= "protein required for nitrogenase activity"  
FT complement (439923..441032)  
FT *tag= r  
FT /standard_name= "ORF K18"  
FT /gene= "fixB"  
FT /product= "protein required for nitrogenase activity"  
FT complement (441042..441899)  
FT *tag= s  
FT /standard_name= "ORF K19"  
FT /gene= "fixA"  
FT /product= "protein required for nitrogenase activity"  
FT complement (442316..442636)  
FT *tag= t  
FT /standard_name= "ORF K20"  
FT /product= "protein of unknown function"  
FT complement (443313..443879)  
FT *tag= u  
FT /standard_name= "ORF K21"  
FT /product= "protein of unknown function"  
FT 444337..445029  
FT *tag= v  
FT /standard_name= "ORF K22"  
FT /product= "ferrodoxin-like protein"  
FT /note= "homologous to the NifQ gene"  
FT 445088..446602  
FT *tag= w  
FT /standard_name= "ORF K23"  
FT /gene= "dctA"  
FT /product= "C4-dicarboxylate transport protein"  
FT /note= "homologous to the DctAI gene"  
FT 446599..447843  
FT *tag= x  
FT /standard_name= "ORF L1"  
FT /product= "cytochrome P450-like protein"  
FT /note= "homologous to the CamC gene"  
FT 447844..448500  
FT *tag= y
```

```
FT /standard_name= "ORF L2"  
FT /product= "gamma-hexachlorocyclohexane-dechlorinase-like  
FT protein"  
FT /note= "homologous to the LinA gene"  
FT 448497..450203  
FT *tag= z  
FT /standard_name= "ORF L3"  
FT /product= "putative protein with degradative function"  
FT 450341..451396  
FT *tag= aa  
FT /standard_name= "ORF L4"  
FT /product= "luciferase alpha-subunit-like protein"  
FT /note= "homologous to the LuxA gene"  
FT 452980..454494  
FT *tag= ab  
FT /standard_name= "ORF L6"  
FT /gene= "nifD"  
FT /product= "alpha-subunit of FeMo protein of nitrogenase"  
FT 454590..456131  
FT *tag= ac  
FT /standard_name= "ORF L7"  
FT /gene= "nifK"  
FT /product= "beta-subunit of FeMo protein of nitrogenase"  
FT 456187..457677  
FT *tag= ad  
FT /standard_name= "ORF L8"  
FT /product= "protein involved in FeMo co-factor  
FT biosynthesis"  
FT /note= "homologous to the NifB gene"  
FT 457887..459096  
FT *tag= ae  
FT /standard_name= "ORF L9"  
FT /product= "protein involved in FeMo co-factor  
FT biosynthesis"  
FT /note= "homologous to the FixF gene"  
FT 459093..459575  
FT *tag= af  
FT /standard_name= "ORF L10"  
FT /product= "protein of unknown function"  
FT /note= "homologous to the NifX gene"  
FT 459579..460067  
FT *tag= ag  
FT /standard_name= "ORF L11"  
FT /product= "protein of unknown function"  
FT 460501..460920  
FT *tag= ah  
FT /standard_name= "ORF L12"  
FT /product= "protein similar to part of the Fe protein  
FT of nitrogenase"  
FT /note= "homologous to the NifH gene"  
FT 461228..461545  
FT *tag= ai  
FT /standard_name= "ORF L13"  
FT /product= "protein of unknown function"  
FT 463201..464739  
FT *tag= aj  
FT /standard_name= "ORF L14"  
FT /product= "peptidase-like protein"  
FT /note= "homologous to the bi-MPP gene"  
FT 464736..466079  
FT *tag= ak  
FT /standard_name= "ORF L15"  
FT /product= "processing protease-like protein"  
FT /note= "homologous to the PP gene"  
FT 466590..467021  
FT
```

```
Query Match 28.3%; Score 265; DB 19; Length 534720;  
Best Local Similarity 56.6%; Pred. No. 3e-73;  
Matches 512; Conservative 0; Mismatches 390; Indels 3; Gaps 1;  
QY 22 ATCTTCGTCGGGTCATCGTGGTTGGTTGGATCTGCATGTCGCGAAGTTGAGNA 81  
DB 6087 ATTTGGTTCGAGGACACAAAGGATATGTCGCGAGCCGCAATAATTGATCGTTCCTCC 6028
```

QY 82 CAAGGTTTACCAATCTCGTCTTAAACACACGCGAGCTTGTATCTCACTCGTCAAGCC 141  
 DB 6027 GAGGATTTGGAAGTCACTGTTGCAGATAGCA---AAAGCTTGAATCTGACGGCGCAAG 5971  
 QY 142 GATGTTGAATCTCTTTTCTCAAGAGAACGAGTTTATGTAACTCTAGCAGCAGCTAAA 201  
 DB 5970 GAAGTTGAGAAATTTCTATTAAAGGAAAGCCGACGCGGTCTAATATGCGCGCAGGA 5911  
 QY 202 GTTGGTGGTATTCACGCTAACACACCTATCTCTGCTGATTTTCATTTGGTGTCAATCCAG 261  
 DB 5910 GTGGTGGGATCTCGCAATGATATATGCGCGCTGACTTCTATCAAAACCTTATC 5851  
 QY 262 ATTACACCAATCTGATCACTCTGCAATGACACGCTGTGAAGAGCTTCTCTCTCTT 321  
 DB 5850 ATGAGGCTTATGCTATGAGGCTCTTCCGAGTGGGCTTGAAGAGCTTCTCTCTCTT 5791  
 QY 322 GGATCATCTGCAATTTACCTAAATTTGCTCTGACGCAATTTCTGAGTCTCTTTGTTA 381  
 DB 5790 GGATCGAGTTGCATATATCGAAGTATGCGGCGAGCCCATTAAGGAGAGGCTCTATTA 5731  
 QY 382 ACAGCATCTGTAACCACTAATGAGTGTATGCTATTTGCTTAAGTCTGCTGGATTAA 441  
 DB 5730 ACCGGACCACTTGAGCGGACCAACGAGTGTATGCGATCGCAAAATCGCGGCAATTA 5671  
 QY 442 ACTTGTGAGGCTTATAGGATTCAGACGCAATGCGATGCAATCTCTGCGATGCTACTAAT 501  
 DB 5670 TTGTGTCAAGCTGATCTGTAAGCAATACGGCGCAACTTCAATACGCAATGCGCAAT 5611  
 QY 502 CTCTATGTTGCTTAATGACAAATTTCCACCGGAGAAATTTCTCATGTGCTCTCTCTTATG 561  
 DB 5610 CTCTATGTTGCTTAATGACAAATTTCTCATGTGCTCTCTCTCTTATG 5551  
 QY 562 AGGAGTTTCCAGGCGGAAAGTGAATGAGCGGAGGAGTTGTGTTGGTGTGGGTCAGGT 621  
 DB 5550 CGCAAGACATGAGGCAAGATTAAGACCTTTGGTGTCTTATATATGCGGAGCGGC 5491  
 QY 622 AGTCCGTTGAGGAGTTCTTGATGTTGATGTTGGCTGATGCTTGTGTTTCTTGTG 681  
 DB 5490 ACACCTACTCGAGACTTTTGTACAGTGAAGTCTCTCGAGCGCTCTCTCTCTT 5431  
 QY 682 GATCGATACGCGGGTTGAGCAGTGTAAATTTGGAAGTGTGTCAGAAAGTGAATTAAGA 741  
 DB 5430 AAGCATTATTCGAAACGGAACACATTAACATAGGCTCGCGGGGGAATAAGTATCAT 5371  
 QY 742 GAGTTGGCTGAGTTGTGTAAGGTTGTGTTGTTTGAAGGAGGCTTGGATGGATTCG 801  
 DB 5370 GAACTAGCCCAATCTGTCGCGGTGTTGTTGTTTAAAGGCGATATAGTCTTCGACACA 5311  
 QY 802 ACTAAGCCAGATGGCACCGAGGAACTTATGGAAGCTCAAGCTCGCGTCTTGGT 861  
 DB 5310 TCCNAGCCGAGCGGAACCGCAAGCTTTTATCTAGCGAAGAGCTCTGTGCGATGGT 5251  
 QY 862 TGGACACCTAAGGTTTCTCTTAGAGATGCTGAGCCAACTTATGATTTGATTTGAAG 921  
 DB 5250 TGGCGGCCGAGACCTCGCTCGAGCTGGGACTGGCCAAATCTATGAATCGTTTGTGACG 5191  
 QY 922 AATGT 926  
 DB 5190 AATGT 5186

RESULT 15  
 AAV30459/c  
 ID AAV30459 standard; DNA; 536165 BP.  
 XX  
 AC AAV30459;  
 XX  
 DT 06-JUL-1999 (first entry)  
 XX  
 DE Rhizobium species symbiotic plasmid pNGR234.  
 XX  
 KW Symbiosis; open reading frame; ORF; plasmid; vector; transportation;

KW degradation; metabolism; host range; nitrogen fixation; nodulation;  
 KW legume; plant; ds.  
 XX Rhizobium sp.  
 OS  
 PN WO9802560-A2.  
 XX  
 PD 22-JAN-1998.  
 XX  
 PF 10-JUL-1997; 97WO-1B00950.  
 XX  
 PR 20-MAY-1997; 97GB-0010395.  
 PR 12-JUL-1996; 96EP-0730001.  
 XX  
 PA (MOLE-) INST MOLECULAR BIOTECHNOLOGY.  
 PA (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.  
 XX  
 PI Broughton WJ, Freiberg CB, Perret XP, Rosenthal A;  
 XX WPI; 1998-110606/10.  
 XX  
 PT New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to  
 PT develop products for modifying plant characteristics, e.g. nitrogen  
 PT fixation, synthesis of compounds and stress response  
 XX  
 PS Claim 1; Fig 3; 228pp; English.  
 XX  
 CC This is the nucleotide sequence of the plasmid pNGR234a isolated from  
 CC Rhizobium sp. NGR234. Open reading frames (ORF) derivable from the  
 CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can  
 CC be used e.g. in the transportation of compounds to and from an organism  
 CC which is a host to at least one of the nucleotide sequences, ORFs or  
 CC proteins, the degradation and/or metabolism of organic, inorganic,  
 CC natural or xenobiotic substances in a host organism or the modification  
 CC of the host range, nitrogen fixation abilities; for obtaining a synthetic  
 CC minimal set of ORFs required for functional Rhizobium-legume symbiosis,  
 CC especially for nodulation efficiency on host plants.  
 XX  
 SQ Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;  
 Query Match 28.3%; Score 265; DB 19; Length 536165;  
 Best Local Similarity 56.6%; Pred. No. 3e-73;  
 Matches 512; Conservative 0; Mismatches 390; Indels 3; Gaps 1;  
 QY 22 ATCTTCGTCGGGTGTCATCGTGTGTTGGTTCGATCTGCGCATTCGCGCAAGCTTCAGGAA 81  
 DB 6087 ATTTGGTTCGAGACACAAAGGATATGTCGCGAGCGCATTAATTCGATCGTTCGCTCC 6028  
 QY 82 CAAGGTTTTCACCAATCTCGTCTTTAAACACACGCGAGCTTGTATCTCACTCGTCAAGCC 141  
 DB 6027 GAGGATTCGGAAGTCACTGTTGCAGATAGCA---AAAGCTTGAATCTGACGGCGCAAG 5971  
 QY 142 GATGTTGAATCTCTTTTCTCAAGAGAACGAGTTTATGTAACTCTAGCAGCAGCTAAA 201  
 DB 5970 GAAGTTGAGAAATTTCTATTAAAGGAAAGCCGACGCGGTCTAATATGCGCGCAGGA 5911  
 QY 202 GTTGGTGGTATTCACGCTAACACACCTATCTCTGCTGATTTTCATTTGGTGTCAATCTCCAG 261  
 DB 5910 GTGGTGGGATCTCGCAATGATATGATGCGCGTCTGATCTCTATCAATCAAAACCTTATC 5851  
 QY 262 ATTACACCAATGTCATCTCTGATATGAGCAGCGGTGTGAAGAGCTTCTCTCTCTT 321  
 DB 5850 ATGAGGCTTATGTCATTTGAGGCTCTCTTCGAGTGGGCTTGAAGAGCTTCTCTCTT 5791  
 QY 322 GGATCATCTGATTTTACCTTAAATTTGCTCTGAGCCAAATTCCTGAGTCTGCTTTGTTA 381  
 DB 5790 GGATCGAGTTGCTATATCCGAAGTATGCGGCGAGCCCATTAAGGAGAGGCTCTATTA 5731  
 QY 382 ACAGCATCTGTAACCACTAATGAGTGTATGCTATTTGCTTAAGTCTGCTGGATTAA 441  
 DB 5730 ACCGGACCACTTGAGCGGACCAACGAGTGTATGCGATCGCAAAATCGCGGCAATTA 5671  
 QY 442 ACTTGTGAGGCTTATAGGATTCAGACGCAATGCGATGCAATCTCTGCGATGCTACTAAT 501

Db	5670	TTGTGTCAAGCGTATCGTAAGCAATACGGCGAACTTCATATACCCATGCCGCAAAAT	5611
Qy	502	CTCTATGTCCTAATGACAAATTTCCACCGGAGAAATTTCTCATGTGCTTCTGCTCTTATG	561
Db	5610	CTCTATGGCCACGGGATAGTTGGATCTTAATCCAGCCACGTGCTCCCTGCCTTAATA	5551
Qy	562	AGGAGTTCCACGAGCGGAAAGTGAATGGAGCGGAGGAAGTTGTGGTGTGGGGTACAGGT	621
Db	5550	CGCAAAAGCACATGAGGCAAGATTTAAAGACCTTGGGTGCTTGTCTATATGGGGAAGCGGC	5491
Qy	622	AGTCGGTTGAGGAGTTCTTGATGTTGATGTTGGCTGATGCTTGTGCTTTCTTCTGCTG	681
Db	5490	ACACCTACTCGAGACTTTTGTACAGTGAAGACTGCTCCGAGCCCTGGTCTTCTTACTT	5431
Qy	682	GATCGATACAGCGGTTGGAGCATGTTAAACATTGGAAGTGGTCAAGAGTGACTATTAGA	741
Db	5430	AAGCATTTATCCGAACGGACACACATTAAACATAGGCTCCGGGGGGAAATAGTATCATT	5371
Qy	742	GAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGGAAAGCTTCGATGGGATTCG	801
Db	5370	GAACTAGCCACATCGTCTCGCGTGTGTTGGTTTAAAGGGCATATAGTCTTCGACACA	5311
Qy	802	ACTAGCCAGATGGCACACGGAGAACTTATGCACAGCTCAAAGCTCGGCTCTTGGGT	861
Db	5310	TCAAGCCGGACGGAAACGCCAAGAAAGCTTTATCTAGCGAAAGACTCGTTCGATGGGT	5251
Qy	862	TGGACACCTAAGGTTTCTCTTAGAGATGGTCTGAGCCAAACTTATGATTGGTATTGAAG	921
Db	5250	TGGGCCCCGAGACCTCGCTCGAGCTGGGACTGGCCAAATCCTATGAATCGTTTGTACGC	5191
Qy	922	AATGT 926	
Db	5190	AATGT 5186	

Search completed: June 3, 2003, 02:47:57  
 Job time : 300 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2003, 02:38:25 ; Search time 2637 Seconds  
(without alignments)  
10330.006 Million cell updates/sec

Title: US-10-089-014-2

Perfect score: 936

Sequence: 1 atgtgtgacaaatctgccaa.....tgagaagtgtttgcaacoga 936

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	936	100.0	939	8	AB034806	AB034806 Arabidops
2	936	100.0	939	8	AF045286	AF045286 Arabidops
3	936	100.0	70836	8	AC010556	AC010556 Arabidops
4	667.8	71.3	1018	8	AY096364	AY096364 Arabidops
5	667.8	71.3	1341	8	AY063880	AY063880 Arabidops
6	667.8	71.3	65262	2	AC025908	AC025908 Arabidops
7	667.8	71.3	75133	8	F2H15	AC034106 Sequence
8	659	70.4	1353	8	AY089031	AY089031 Arabidops
9	430.6	46.0	130393	2	AP003565	AP003565 Oryza sat
10	391	41.8	1490	6	AX172657	AX172657 Sequence
11	354	37.8	10029	1	AE013290	AE013290 Methanosa
12	301.8	32.2	8458	1	AE007718	AE007718 Clostridi
13	297	31.7	1020	6	AF03794	AF03794 Sequence 15
14	297	31.7	4435	1	MAP223833	AJ223833 Mycobacte
15	297	31.7	4435	6	AF03783	AF03783 Sequence 4
16	295.4	31.6	1020	6	AF03792	AF03792 Sequence 13
17	295.4	31.6	7995	6	AF03782	AF03782 Sequence 3
18	295.4	31.6	8938	1	MAS223832	AJ223832 Mycobacte
19	295.4	31.6	25861	1	AF125999	AF125999 Mycobacte
20	295.4	31.6	71286	1	AF143772	AF143772 Mycobacte
21	285.4	30.5	130393	2	AP003565	AP003565 Oryza sat
22	284.6	30.4	10399	1	AE013711	AE013711 Yersinia
23	284.6	30.4	20588	1	YPS251712	AJ251712 Yersinia
24	279.4	29.9	26913	1	YEU46859	U46859 Yersinia en
25	276.8	29.6	19553	1	AE007023	AE007023 Mycobacte
26	276.4	29.5	330743	14	PRU42580	U42580 Paramecium
27	275.2	29.4	12732	6	AR128139	AR128139 Sequence
28	275.2	29.4	37316	1	MSGY456	AD000001 Mycobacte
29	275.2	29.4	38300	1	MTCY277	Z79701 Mycobacteri
30	274.2	29.3	20745	1	YPE251713	AJ251713 Yersinia
31	274.2	29.3	205050	1	AJ414155	AJ414155 Yersinia
32	271.8	29.0	20148	1	AE008793	AE008793 Salmonell
33	271.8	29.0	250050	1	AL627273	AL627273 Salmonell
34	270.2	28.9	12016	1	AE009407	AE009407 Agrobacte
35	270.2	28.9	12862	1	AE008206	AE008206 Agrobacte
36	269.8	28.8	12666	1	AE005429	AE005429 Escherich
37	269.8	28.8	14002	1	AF061251	AF061251 Escherici
38	269.8	28.8	14029	6	AX029262	AX029262 Sequence
39	269.8	28.8	14187	6	AR204225	AR204225 Sequence
40	269.8	28.8	31482	1	AB008676	AB008676 Escherich
41	269.8	28.8	286857	1	AP002559	AP002559 Escherich
42	268.6	28.7	12897	1	AF285084S1	AF285084 Salmonell
43	265	28.3	11429	1	AE000064	AE000064 Rhizobium
44	265	28.3	320040	6	A79350	A79350 Sequence 1
45	265	28.3	320040	6	A93002	A93002 Sequence 1

# ALIGNMENTS

RESULT 1	AB034806	939 bp	mrna	linear	PLN 13-MAY-2000
LOCUS	Arabidopsis thaliana AtFX mRNA for				
DEFINITION	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase, complete cds.				
ACCESSION	AB034806				
VERSION	AB034806.1	GI:7804473			
KEYWORDS	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase.				
SOURCE	Arabidopsis thaliana	CDNA to mRNA.			
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.				
REFERENCE	1 (sites)				

AUTHORS Nakayama, K., Maeda, Y., Wang, X. and Jigami, Y.  
 TITLE Expression of the genes for GDP-fucose synthesis in yeast  
 JOURNAL Saccharomyces cerevisiae  
 REFERENCE Unpublished  
 2 (bases 1 to 939)  
 AUTHORS Nakayama, K. and Jigami, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-NOV-1999) Ken-ichi Nakayama, National Institute of  
 Bioscience and Human Technology, Molecular Biology Department, 1-1  
 Higashi, Tsukuba, Ibaraki 305-8566, Japan  
 (E-mail:nakayama@nibh.go.jp, Tel:81-298-54-6226,  
 Fax:81-298-54-6220)  
 FEATURES Location/Qualifiers  
 source 1..939  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 gene 1..939  
 /genes="AtFX"  
 CDS 1..939  
 /genes="AtFX"  
 /codon\_start=1  
 /product="GDP-4-keto-6-deoxy-D-mannose-3,  
 5-epimerase-4-reductase"  
 /protein\_id="BAA95670.1"  
 /db\_xref="GI:7804474"  
 /translation="MSDKSAKIPVAGHRLVGSIVRKLQEQFQTNLVKTHAELDLT  
 ROADVSFQSEKPVVILAAKVGIGHANNTVPADFIGVNIQIOTNLVHSAVEHGVK  
 KLLFLGSSCIYKFAQPIPESEALLTASLEPNEWYAIKAGIKTCQAYRLOHGND  
 ISGMPTNLGPNDFPNSHVLPAHFAKNGVGEVVMVGTSPLRFLHVD  
 LADACVFLDRYGLHVNIGSGEVTIRELAELVKEVVGFEKGLGWDCTKPDGTPRK  
 LMDSSKLASLIGWTPKVSRLDGLSQTYDYLKNCNR"  
 BASE COUNT 229 a 190 c 240 g 280 t

Query Match 100.0%; Score 936; DB 8; Length 939;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-265;  
 Matches 936; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTGGATCTGCC 60  
 DB 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTGGATCTGCC 60  
 QY 61 ATTGTCGGCAAGCTTCAGGAACAAGTTTCCAAATCTCGTTCTTAAACACACGCCGAG 120  
 DB 61 ATTGTCGGCAAGCTTCAGGAACAAGTTTCCAAATCTCGTTCTTAAACACACGCCGAG 120  
 QY 121 CTGTGCTCACTGCTCAAGCCGATGTTGAATCTCTTTTCAAGAGAAGCCAGTTTAT 180  
 DB 121 CTGTGCTCACTGCTCAAGCCGATGTTGAATCTCTTTTCAAGAGAAGCCAGTTTAT 180  
 QY 181 GTAACTCTAGCAGCAGTAAAGTTGGTGGTATTTCAGCTAACACACCTATCTGCTGAT 240  
 DB 181 GTAACTCTAGCAGCAGTAAAGTTGGTGGTATTTCAGCTAACACACCTATCTGCTGAT 240  
 QY 241 TTCAATGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTTCGATATGACGACGCT 300  
 DB 241 TTCAATGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTTCGATATGACGACGCT 300  
 QY 301 GTGAAGAAGCTTCTCTCTTGGATCATCTGATATTAACCTAAATTTGCTCTCAGCCA 360  
 DB 301 GTGAAGAAGCTTCTCTCTTGGATCATCTGATATTAACCTAAATTTGCTCTCAGCCA 360  
 QY 361 ATTCCTGAGTCTCTTTGTTTAACAGCATCGTTTGAACCAACTAAATGAGTGGTATGATT 420  
 DB 361 ATTCCTGAGTCTCTTTGTTTAACAGCATCGTTTGAACCAACTAAATGAGTGGTATGATT 420  
 QY 421 GCTAAGATCGTGGGATTAAAGCTTCTCAGGCTTATAGGATTCAGACGATGGGATGCA 480  
 DB 421 GCTAAGATCGTGGGATTAAAGCTTCTCAGGCTTATAGGATTCAGACGATGGGATGCA 480  
 QY 481 ATCTCTGGCATGCCTACTAATCTCTATGCTCTTAATGACAAATTTCCACCGGAGATTCT 540  
 DB 481 ATCTCTGGCATGCCTACTAATCTCTATGCTCTTAATGACAAATTTCCACCGGAGATTCT 540

QY 541 CATGTGCTTCCTGCTCTTATGAGGAGGTTCCACGAGGCGAAAGTGAATCGAGCGGAGAA 600  
 DB 541 CATGTGCTTCCTGCTCTTATGAGGAGGTTCCACGAGGCGAAAGTGAATCGAGCGGAGAA 600  
 QY 601 GTTGTGGTGGGTACAGGTAGTCCGTTGAGGGAGTTCTTCATGTTGATGATTGGCT 660  
 DB 601 GTTGTGGTGGGTACAGGTAGTCCGTTGAGGGAGTTCTTCATGTTGATGATTGGCT 660  
 QY 661 GATGCTTGTTTCTTCTGCTGGATCGATACAGCGGGTTGGAGCATGTTAAACATTGGAAT 720  
 DB 661 GATGCTTGTTTCTTCTGCTGGATCGATACAGCGGGTTGGAGCATGTTAAACATTGGAAT 720  
 QY 721 GGTCAAGAAGTCACTATTAGAGAGTTGGTGAAGTTGGTGAAGAGAGTTGTTGTTTGA 780  
 DB 721 GGTCAAGAAGTCACTATTAGAGAGTTGGTGAAGTTGGTGAAGAGAGTTGTTGTTTGA 780  
 QY 781 GGGAAAGCTTGGATGGATTGCTAAGCCAGATGTCACACCGAGGAAACTTATGGACAGC 840  
 DB 781 GGGAAAGCTTGGATGGATTGCTAAGCCAGATGTCACACCGAGGAAACTTATGGACAGC 840  
 QY 841 TCAGAAGCTCGCTCTTGGTGGTGGACACCTAGAGTTCTCTTAGAGATGGTCTGAGCCAA 900  
 DB 841 TCAGAAGCTCGCTCTTGGTGGTGGACACCTAGAGTTCTCTTAGAGATGGTCTGAGCCAA 900  
 QY 901 ACTTATGATTGGTATTGGAAGAAATGTTGCAACCGA 936  
 DB 901 ACTTATGATTGGTATTGGAAGAAATGTTGCAACCGA 936  
 RESULT 2  
 AF045286  
 LOCUS Arabidopsis thaliana 939 bp DNA linear PLN 19-APR-2000  
 DEFINITION GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase (GER1) gene,  
 complete cds.  
 ACCESSION AF045286  
 VERSION AF045286.3 GI:7596994  
 KEYWORDS Arabidopsis thaliana.  
 SOURCE Arabidopsis thaliana.  
 ORGANISM Arabidopsis thaliana.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 939)  
 AUTHORS Bonin, C.P. and Reiter, W.D.  
 TITLE A bifunctional epimerase-reductase acts downstream of the MUR1 gene  
 product and completes the de novo synthesis of GDP-L-fucose in  
 Arabidopsis  
 JOURNAL Plant J. 21 (5), 445-454 (2000)  
 MEDLINE 20223138  
 PUBMED 10758496  
 REFERENCE 2 (bases 1 to 936)  
 AUTHORS Bonin, C.P., Potter, I., Vanzin, G.F. and Reiter, W.-D.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JAN-1998) Molecular and Cell Biology, University of  
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA  
 REFERENCE 3 (bases 1 to 936)  
 AUTHORS Bonin, C.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-OCT-1999) Molecular and Cell Biology, University of  
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA  
 REFERENCE 4 (bases 1 to 939)  
 AUTHORS Reiter, W.-D.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-APR-2000) Molecular and Cell Biology, University of  
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA  
 REFERENCE 5 (bases 1 to 939)  
 AUTHORS Reiter, W.-D.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-APR-2000) Molecular and Cell Biology, University of  
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA  
 REMARK Sequence update by submitter  
 COMMENT On Apr 19, 2000 this sequence version replaced gi:6016478.  
 FEATURES Location/Qualifiers  
 source 1..939  
 /organism="Arabidopsis thaliana"



```

/cultivar="Columbia ecotype"
/db_xref="taxon:3702"
<1..>939
/gene="GER1"
<1..>939
/gene="GER1"
/product="GDP-4-keto-6-deoxy-D-mannose-3,
5-epimerase-4-reductase"
1..939
/gene="GER1"
/feature="enzyme in pathway for de novo synthesis of
GDP-L-fucose"
/codon_start=1
/product="GDP-4-keto-6-deoxy-D-mannose-3,
5-epimerase-4-reductase"
/protein_id="AAC02703.2"
/db_xref="GI:6016479"
/translation="MSDSAKIFVAGHGLVGSATVRKLOEGFTNLVLTAEIDLIT
RADVESFSPQKPVYVILAAKVGCIHANNTYPADFIGVNLQIQTNVHSAIEHGVK
KULFGSSCIYKFPAPQIPESALITLSEPTNEWYAIAGIKTCQAYRIQHGWD
ISGMPNLNYPNDHPENSHVLPALMRPEAKVNGAEVVGWGTGSPLEFLHVD
LADACVFLDRLYSGLEHVNIGSQEVTIRELAELVKEVVGFEGLGWDCTKPDGTPRK
LMDSSKLSLGTWPKVSLRDLGSOTVDWYLNKVCNR"
BASE COUNT 229 a 190 c 240 g 280 t
ORIGIN
Query Match 100.0%; Score 936; DB 8; Length 939;
Best Local Similarity 100.0%; Pred. No. 1.2e-265;
Matches 936; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTGACAAATCTCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 60
DB 1 ATGCTGACAAATCTCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 60
QY 61 ATTGTCGGAAGCTTCAGGAACAAGTTTCAACAATCTCGTCTTAAACACACGCCGAG 120
DB 61 ATTGTCGGAAGCTTCAGGAACAAGTTTCAACAATCTCGTCTTAAACACACGCCGAG 120
QY 121 CTGATCTCACTCGTCAAGCCGATGTGAATCTCTTTTCAAGGAAGCCAGTTTAT 180
DB 121 CTGATCTCACTCGTCAAGCCGATGTGAATCTCTTTTCAAGGAAGCCAGTTTAT 180
QY 181 GTAATCTAGCAGCAGTAAAGTTGGTGTATTCAGCTTAACACACCTATCTGCTGAT 240
DB 181 GTAATCTAGCAGCAGTAAAGTTGGTGTATTCAGCTTAACACACCTATCTGCTGAT 240
QY 241 TTCAATGGTGTCAATCTCCAGATTCAGCAATGTGATCCACTCTCATATGAGCAGGT 300
DB 241 TTCAATGGTGTCAATCTCCAGATTCAGCAATGTGATCCACTCTCATATGAGCAGGT 300
QY 301 GTGAAGAAGCTTCTCTTCTTGGATCATCTGCAATTAACCTAAATTTGCTCTCAGCCA 360
DB 301 GTGAAGAAGCTTCTCTTCTTGGATCATCTGCAATTAACCTAAATTTGCTCTCAGCCA 360
QY 361 ATTCCTGAGTCTGCTTTGTTAAACAGATCTTGAACCACTAATGAGTGGTATGCTATT 420
DB 361 ATTCCTGAGTCTGCTTTGTTAAACAGATCTTGAACCACTAATGAGTGGTATGCTATT 420
QY 421 GCTAAGATCGCTGGGATTAAGCTTCTCAGGCTTATAGGATTCAGCAGCGGATGCA 480
DB 421 GCTAAGATCGCTGGGATTAAGCTTCTCAGGCTTATAGGATTCAGCAGCGGATGCA 480
QY 481 ATCTCTGGCATGCTACTAATCTCTATGCTCTAATGACAATTTCCACCGGAGATTCT 540
DB 481 ATCTCTGGCATGCTACTAATCTCTATGCTCTAATGACAATTTCCACCGGAGATTCT 540
QY 541 CATGTGCTTCTCTCTTATGAGGATTCACAGGCGAAGTGAATGAGCGGAGGAA 600
DB 541 CATGTGCTTCTCTCTTATGAGGATTCACAGGCGAAGTGAATGAGCGGAGGAA 600
QY 601 GTTGTGGTGGGATACAGTACTCGGTTGAGGAGTCTTTCATCTGATGATTTGGCT 660
DB 601 GTTGTGGTGGGATACAGTACTCGGTTGAGGAGTCTTTCATCTGATGATTTGGCT 660

```

```

QY 661 GATGCTTGTTTCTTGCTGATCGATACACGGGTTGGAGCATGTTAAACATTGGAAGT 720
DB 661 GATGCTTGTTTCTTGCTGATCGATACACGGGTTGGAGCATGTTAAACATTGGAAGT 720
QY 721 GGTCAAGAGTCACTATTAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGTTTGA 780
DB 721 GGTCAAGAGTCACTATTAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGTTTGA 780
QY 781 GCGAAGCTTGGATGGATTCGACTAAGCCAGATGCGACACCGAGGAACTTATGGACAGC 840
DB 781 GCGAAGCTTGGATGGATTCGACTAAGCCAGATGCGACACCGAGGAACTTATGGACAGC 840
QY 841 TCAAGCTCGCTCTTTGGTTGGACACCTAAGTTTCTTATAGATGCTTGAAGCCAA 900
DB 841 TCAAGCTCGCTCTTTGGTTGGACACCTAAGTTTCTTATAGATGCTTGAAGCCAA 900
QY 901 ACTTATGATGCTTATTTGAAGATGTTTGAACCGA 936
DB 901 ACTTATGATGCTTATTTGAAGATGTTTGAACCGA 936

```

```

RESULT 3
AC010556 70836 bp DNA linear PLN 15-JUN-2001
Arabidopsis thaliana chromosome 1 BAC T18K17 genomic sequence,
complete sequence.
AC010556
AC010556.6 GI:12324309
HTG
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
Lin.X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Matti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC T18K17 genomic sequence
Unpublished
2 (bases 1 to 70836)
Lin.X. and Kaul, S.
Direct Submission
Submitted (15-SEP-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 70836)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280860.
Address all correspondence to:at@tigr.org

```

BAC clone T18K17 is from Arabidopsis thaliana chromosome 1. The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), Glimmer (a variant of Glimmer, see Mihaela Pertea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as

hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

source

1. 70836

```

I: : 70850
/organism="Arabidopsis thaliana"

```

[/culturivay="Columbia"/](#)

dbref="3702  
/columbia/paravinc/  
dbref="3702

```
/db_xref="taxon:3702"
/chromosome="11"
```

```
/chromosome="1"
/chromosome="17"
```

```
/clone="T18K17"
```

MKLDVNGTLPVSLVSTYTSWTKASSMIFLDQPVGAGFSYRTQLLNKPSDSEAKRIH EFLQWLGHQBFSSNPFYVGDSDGSMVVPATVQBIISKNGYECNPPINLQGYVLGN PLTFDYDYNLSIFPAHGMALISDELFSKTKCKGDRYRNVHPRTECLKLFEEYVNC TNSICORRIIDPFCEPTNCVIRFLAAYWANDETVRKALQIKETIGEVWRCHYG IPYNDIKSIPYHMNSINGVRSLSIYSGDHPFVFLGTQAWIRSLNYSVIDDNRPW									
Query Match	100.0%	Score	936;	DB	8;	Length	70836;		
Best Local Similarity	100.0%	Pred. No.	2.1e-265;						
Matches	936;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ATGTCGTGACAAATCTGCCAAATCTTCGTCGGGGTCACTGCTGGTGTGGTGGATCTGCC	60						
Db	21556	ATGTCGTGACAAATCTGCCAAATCTTCGTCGGGGTCACTGCTGGTGTGGTGGATCTGCC	21615						
QY	61	ATTGTCGCGAAGCTTCAGAAACAAGTTTTCACAAATCTCGTTCTTTAAAAACACACGCCGAG	120						
Db	21616	ATTGTCGCGAAGCTTCAGAAACAAGTTTTCACAAATCTCGTTCTTTAAAAACACACGCCGAG	21675						
QY	121	CTTGATCTCACTCGTCAAGCCGATGTGAATCTCTTCTTCTCAAGAGACCGATTAT	180						
Db	21676	CTTGATCTCACTCGTCAAGCCGATGTGAATCTCTTCTTCTCAAGAGACCGATTAT	21735						
QY	181	GTAAATCCTAGCAGCAGTAAAGTTGTGTGTTATTCAGCTAAACACACTATCTCTGCTGAT	240						
Db	21736	GTAAATCCTAGCAGCAGTAAAGTTGTGTGTTATTCAGCTAAACACACTATCTCTGCTGAT	21795						
QY	241	TTCAATGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACGACGGT	300						
Db	21796	TTCAATGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACGACGGT	21855						
QY	301	GTGAGAAAGCTTCTTCTTGTGATCATCTGATTTACCTTAATTTGCTCTCTAGCCA	360						
Db	21856	GTGAGAAAGCTTCTTCTTGTGATCATCTGATTTACCTTAATTTGCTCTCTAGCCA	21915						
QY	361	ATTCTGAGTCTGCTTGTGTTTAAACAGCATCGTCTGACCAACTAAAGAGTGTATCTAT	420						
Db	21916	ATTCTGAGTCTGCTTGTGTTTAAACAGCATCGTCTGACCAACTAAAGAGTGTATCTAT	21975						
QY	421	GCTAAGATCGCTGGGATTAAGACTTGTTCAGGCTTATAGGATTCAGACGGATGGATGCA	480						
Db	21976	GCTAAGATCGCTGGGATTAAGACTTGTTCAGGCTTATAGGATTCAGACGGATGGATGCA	22035						
QY	481	ATCTCTGGATGCTACTAATCTCTATGCTCTATATGACAAATTTCCACCGGAGATTCT	540						
Db	22036	ATCTCTGGATGCTACTAATCTCTATGCTCTATATGACAAATTTCCACCGGAGATTCT	22095						
QY	541	CATGCTCTTCTGCTCTTATGAGGAGTTCCAGGCGCAAGTGAATGAGCGGAGAA	600						
Db	22096	CATGCTCTTCTGCTCTTATGAGGAGTTCCAGGCGCAAGTGAATGAGCGGAGAA	22155						
QY	601	GTGTGCTGTGGGTCACAGTATCGCTTTCAGGAGTCTTTCATGATTTGGCT	660						
Db	22156	GTGTGCTGTGGGTCACAGTATCGCTTTCAGGAGTCTTTCATGATTTGGCT	22215						
QY	661	GATGCTGTGTTTCTTGTGATCATACAGCGGTTGGAGATTTAACTTGAAT	720						
Db	22216	GATGCTGTGTTTCTTGTGATCATACAGCGGTTGGAGATTTAACTTGAAT	22275						
QY	721	GGTCAAGAGTGAATTAAGAGTTGGTGGTGGTCAAGAGGTTGCTGTTTGA	780						
Db	22276	GGTCAAGAGTGAATTAAGAGTTGGTGGTGGTCAAGAGGTTGCTGTTTGA	22335						
QY	781	GGGAGCTTGGATGGATTCATAGCCAGATGGCACACCGAGGAACTTATGACAGC	840						
Db	22336	GGGAGCTTGGATGGATTCATAGCCAGATGGCACACCGAGGAACTTATGACAGC	22395						
QY	841	TCAAAGCTCGCTTTTGGTGGACACTAAGTTTCTTTAGAGATGCTCTGAGCCAA	900						
Db	22396	TCAAAGCTCGCTTTTGGTGGACACTAAGTTTCTTTAGAGATGCTCTGAGCCAA	22455						
QY	901	ACTTATGATGTTTGAAGATGTTTTCACCCGA	936						
Db	22456	ACTTATGATGTTTGAAGATGTTTTCACCCGA	22491						

RESULT 4  
AY096364  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AY096364 1018 bp mRNA linear PLN 06-MAY-2002  
Arabidopsis thaliana putative GDP-L-fucose synthetase (At1g17890)  
mRNA, complete cds.  
AY096364  
AY096364.1 GI:20465282  
FLI CDNA.  
Arabidopsis thaliana.  
Arabidopsis thaliana.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis;  
1 (bases 1 to 1018)  
Yamada, K., Banh, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M.,  
Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,  
Tang, C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,  
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,  
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,  
Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,  
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,  
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
Arabidopsis Open Reading Frame (ORF) Clones

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

FEATURES  
Source

1. 1018  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="U18093"  
/note="This clone is in puni 51."  
ecotype: Columbia  
1. 1018  
/gene="At1g17890"  
/gene="At1g17890"  
/codon\_start=1  
/evidence=experimental  
/product="putative GDP-L-fucose synthetase"  
/protein\_id="AA20005.1"  
/db\_xref="GI:20465283"

gene

CDS

The Salk, Stanford, PGEN (SSP) Consortium members constructed and  
sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K.,  
Banh, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M.,  
Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C.,  
Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H.,  
Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,  
Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P.,  
Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.  
Yamada, K. (SSP/PGEN) and Seki, M. (RIKEN GSC) contributed equally  
to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP  
/PGEN) contributed equally to this work as PIs.

/translation="MADNTGSEMKSGSFMLEKSAKIFVAGHRLVGSIVIRKLDQGGF  
TNLVRHSELDITQSDFVESFATEKPVVILAAKVGIIHANNNTYPADFGVNLQI  
QTNVHSAYTHQKLLFLGSSCIYKPAPOPIPEGALLTGLEPNEVWATAKAGI  
KMCQAVRLQHOWDAISGMPNLYGNDNFHPNSHVLPAIMRFFHEKAKKANADEVVW  
GSGSPLEFHLVDLADACVPLMDQYSGFHNVSQVEVTKELAEIYKEVVGFGK  
LWMDITPDGTGPRKMDSSKSLASLWTPKISLKGUSQTYEYILENVVQKKQ"  
988..1018  
/gene="At1g17890"  
3'UTR 277 a 214 c 252 g 275 t  
BASE COUNT 277 a 214 c 252 g 275 t  
ORIGIN  
Query Match 71.3%; Score 667.8; DB 8; Length 1018;  
Best Local Similarity 82.5%; Pred. No. 3e-186;  
Matches 765; Conservative 0; Mismatches 162; Indels 0; Gaps 0;  
QY 1 ATGCTGACAAATCTGCAAAATCTTCGTCGCGGGTCATGCTGTTGGTTGGATCTGCC 60  
DB 43 ATGTTAGAAATCCCGCAAGATCTTGTGCTGGACATAGGGGATGTTGGATCTGCC 102  
QY 61 ATTGTCGCAAGCTTCAGGAACAAGTTTCCACCAATCTCGTTCTTAAACACACGCGGAG 120  
DB 103 ATTGTCGCAAGCTTCAGGATCAGGGTTTCACCAATTTGGTCTTCCGACCCATTCCAG 162  
QY 121 CTTGATCTACTCGTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTAT 180  
DB 163 CTTGATCTACTAGCCAATCAGATGTTGAATCTTTCTTCTACAGAGAAACAGTTTAT 222  
QY 181 GTAATCTTAGCAGCAGTAAAGTTGTTGTTATTCAGCTTAACAACCACTTCTCGCTGAT 240  
DB 223 GTCATCTAGCTGCTGTAAGTTCGGTGGGATTCATGCTTAACAACATACCCAGCTGAT 282  
QY 241 TTCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTCATATGAGCAGCGT 300  
DB 283 TTCATTGGGTCAATCTCCAAATCCAGACCAATGTGATCCACTCTCTTATACACAGCT 342  
QY 301 GTGAAGAAGTTCTTCTCTTGATCATCTGATTTACCTTAATTTGCTCTCAGCCA 360  
DB 343 GTGAAGAAGTCTCTTCTCTTGATCATCTGATTTACCTTAATTTGCTCTCAGCCG 402  
QY 361 ATTCCTGAGTCTCTTGTGTTAAACAGCATCGTTGAACCACTAATCAGTGGTATGCTATT 420  
DB 403 ATTCCTGAGTCTCTGCTTCTTACCGTTCGCTCGAACCGAACCAAGTGGTATGCTATC 462  
QY 421 GCTAAGATCGTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGCGATGGATGCA 480  
DB 463 GCCAAGATCGGGTATTAAAGATGTCTCAAGCTTACCGGCTCAGCATCAGTGGATGCT 522  
QY 481 ATCTCTGGCATGCTACTAATCTCTATGCTCTAATGACAAATTTCCACCGGAGAAATTCCT 540  
DB 523 ATCTCCGGTATGCCACGAATCTGTATGTTGTCAGAAATGATAATTTCCACCGGAGAAATTC 582  
QY 541 CATGCTCTTCTGCTCTTATGAGGAGGTTCCAGCGGCGAAAGTGAATGAGCGGAGGAA 600  
DB 583 CATGCTCTGCTCTCTATGAGGAGGTTTCATGAGCGCAAGCCAAATGATGTCAGCAA 642  
QY 601 GTTGTGGTGGGATACAGTATGCTGCTGAGGAGTCTTGTGATGTTGATGTTGGCT 660  
DB 643 GTTGTGGTGGGAGTGGAGCCGTTGAGGAGTCTTACATGTCGATGTTGGCT 702  
QY 661 GATGCTTGTGTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
DB 703 GATGCTTGTGTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 762  
QY 721 GGTCAAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
DB 763 GGTGTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 822  
QY 781 GGGAGCTTTGGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
DB 823 GGAAGCTTTGTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 882  
QY 841 TCAAGCTTCGCTCTTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 900

883 TCGAAGCTTGCCTCTTGGATGGACACCGAAGATTTGCTTAAAGATGCGCTCTCCCAA 942  
QY 901 ACTTATGATGCTATTTGAAGAATGTT 927  
DB 943 ACTTATGATGCTATTTGAAGAATGTT 969  
RESULT 5  
AY063880 1341 bp mRNA linear PLN 25-APR-2002  
LOCUS Arabidopsis thaliana putative GDP-L-fucose synthetase (At1g17890)  
DEFINITION mRNA, complete cds.  
ACCESSION AY063880  
VERSION AY063880.1 GI:17380847  
KEYWORDS FLI CDNA.  
SOURCE Arabidopsis thaliana.  
ORGANISM Arabidopsis thaliana.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 1341)  
Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B.,  
Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T.,  
Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,  
Davis R.W., Ecker J.R. and Theologis A.  
Arabidopsis Full Length cDNA Clones  
Unpublished  
2 (bases 1 to 1341)  
Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
Lee J.M., Onodera C.S., Quach H.L., Tang C.C., Toriumi M., Wu H.C.,  
Yanamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H.,  
Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,  
Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,  
Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
Shinozaki K., Davis R.W., Ecker J.R. and Theologis A.  
Direct Submission  
Submitted (19-NOV-2001) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
Riken Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN  
Arabidopsis Full-Length cDNA'); Seki M., Narusaka M., Ishida J.,  
Satou M., Kamiya A., Sakurai T., Carninci P., Kawai J.,  
Hayashizaki Y. and Shinozaki K.  
The Salk, Stanford, PGEC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Yamada K., Banh J.,  
Banno F., Chang E., Dale J.M., Goldsmith A.D., Lee J.M.,  
Onodera C.S., Quach H.L., Tang C.C., Toriumi M., Wu H.C.,  
Yanamura Y., Yu G., Yu S., Bowser L., Chen H., Cheuk R., Jones T.,  
Karlin-Neumann G., Kim C., Koesema E., Lam B., Lin J., Meyers M.C.,  
Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A.,  
Davis R.W., Ecker J.R. and Theologis A.  
Yamada K. (SSP/PGEC) and Seki M. (RIKEN GSC) contributed equally to  
this work. Shinozaki K. (RIKEN GSC) and Theologis A. (SSP/PGEC)  
contributed equally to this work as PIs.  
Location/Qualifiers  
1..1341  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="RAFL06-86-B08 (R18093)"  
/note="This clone is in a modified pBluescript vector  
(lambda ZAP) as a XhoI/SstI insert."  
ecotype: Columbia  
1..1341  
/gene="At1g17890"  
1..146  
/gene="At1g17890"  
FEATURES  
source

CDS	147.. 1133 /gene="At1g17890" /codon_start=1 /evidence=experimental /product="putative GDP-L-fucose synthetase" /protein_id="AAL36236.1" /db_xref="GI:17380848"
3'UTR	1134.. 1341 /gene="At1g17890"
BASE COUNT	383 a 275 c 289 g 394 t
ORIGIN	
Query Match	71.3%; Score 667.8; DB 8; Length 1341;
Best Local Similarity	82.5%; Pred. No. 3.le-186;
Matches	765; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
QY	1 ATGTCGTACAATCTGCCAAATCTTCTCGCGGGTCATCTGTTGGTTGGTTCGTC 60
Db	189 ATGTTAGAGAAATCCGCCAAGATCTTTGTCGTGCACATAGGGATTGGTTGATCGCC 248
QY	61 ATTGTCGCAAGCTTCAGGAACAAGTTTCACCAATCTCGTTCCTTAATAACACACGCCGAG 120
Db	249 ATTGTCGCAAGCTCCAGATCAGGGTTTCACCAATTTGGTTCCTTCGNACCATTACAG 308
QY	121 CTTGATCTCATCTCGTCAAGCCGATGTTGAATCTCTTTTCTCAAGAAGAACCCAGTTTAT 180
Db	309 CTTGATCTCACTAGCCAATCAGATCTTGAATCTCTTTTGCTACAGAGAAACCAAGTTTAT 368
QY	181 GTAATCTTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTTACACACACTTCTCTGCTGAT 240
Db	369 GTCATCTAGCTGCTCTAAAGTCGGTGGGATTTCATGCTTAACAACATACCCAGCTGAT 428
QY	241 TTCATTGGTGTCAAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACACGGT 300
Db	429 TTCATTGGGTCNAATCTCCAAATCAGACCAATGTGATCCACTCTGCTTATACACAGGT 488
QY	301 GTGAAGAAGCTTCTCTTCCTTGGATCATCTCGATTTACCTTAATTTGCTCTCCAGCCA 360
Db	489 GTGAAGAAAATCTCTCTCTGTTTCATCTGCTTATACCCCAAAATTTGCTCTCCAGCCG 548
QY	361 ATTCTGAGTCTGCTTTGTTTAACACATCGCTTGNACCAACTAATGATGATGATGCTATT 420
Db	549 ATTCCGGAATCTCGCGCTTCTTACCGGTCGCTCGAACCGACAAACGAATGATGTCGATC 608
QY	421 GCTAAGATCGCTGGGATTAAAGACTTGTGAGGCTTATAGATTTCAGCACGGATGGATGCA 480
Db	609 GCCAAGATCGCGGATTAAAGATGTGTACGCTTACCGGCTTCAGATCATGTGGATGCT 668
QY	481 ATCTCTGGCATCCCTACTAATCTTATGGTCTCTAATGACAAATTTCCACCCGGAGAAATCT 540
Db	669 ATCTCCGGTATGCCACGAATCTGTATGGTCAGAATGATAATTTCCACCCGGAGAAATCC 728
QY	541 CATGTCTTCTGCTCTTATGAGGAGGTTCCACGAGCGAAAGTGAATGAGAGCGGAGAA 600
Db	729 CATGTCTCTGCTCTCTCATGAGGAGGTTTCATGAAGCCAAAGCAATAATGCTGACGAA 788
QY	601 GTTGTGGTGTGGGTACAGTGTAGTCCGTTTGAGGAGTTCTTTCGATGTTGATGATTTGGCT 660
Db	789 GTTGTGGTTTGGGAGTGGGAAGCCGTTTGAGGAGTTTTTACATGTCATGATTTGGCT 848
QY	661 GATGCTTGTGTTTCTTCTGCTGCATCCATACAGCGGTTTGGAGCATGTTTAAACATTGGAAGT 720
Db	849 GATGCTTGTGTTTCTTCTGATGATCAGTACAGCGGATTTGAGCATGTGAATGTAGGAAGT 908
QY	721 GGTCAGAAGTACATATTAGAGATTGGCTGATGTTGGTGAAGAGGTTGTTGTTTGAAG 780
Db	909 GGTGTTGAAGTACCATCAAGAAATTTGGCTGATTTGGTGAAGAGGTTTGGGTTTAAAG 968
QY	781 GCGAAGCTTGGATGGGATTGCACTTAAGCCAGATGGCAACCGAGGAACTTATGGACAGC 840
Db	969 CGAAAGCTTGTGTTGGGATACCAAAACCGAGCGGACACCAAGGAACCTGATGATAGC 1028
QY	841 TCAAGCTCGGCTCTTTGGTGGTGGACACCTTAAGTTTCTCTTGAAGATGCTCTGAGCCAA 900
Db	1029 TCGAAGCTTGGCTCTTTGGGATGACACCGAAGATTTTCGTTAAAGATGCTCTCCCAA 1088
QY	901 ACTTATGATGCTGTTTGAACAAATGTT 927
Db	1089 ACTTATGATGCTTATTTGGAANAATGTT 1115
RESULT 6	
LOCUS	AC025908 65262 bp DNA linear HTG 17-MAR-2000
DEFINITION	Arabidopsis thaliana chromosome 1 clone F20L2, *** SEQUENCING IN PROGRESS ***, 13 unordered pieces.
ACCESSION	AC025908
VERSION	AC025908.1 GI:7259643
KEYWORDS	HTG; HTGS_PHASE1
SOURCE	Arabidopsis thaliana
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
AUTHORS	Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,K., Liu,A., Liu,K., Liu,S., Mukharbek,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Pederspiel,N.A., Theologis,A. and Ecker,J.R.
TITLE	Genomic sequence for Arabidopsis thaliana BAC F20L2 from chromosome I
JOURNAL REFERENCE	Unpublished 2 (bases 1 to 65262)
AUTHORS	Ecker,J.R.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT	* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 2519: contig of 2519 bp in length * gap of unknown length * 2520 contig of 4093 bp in length * gap of unknown length * 6613 contig of 3782 bp in length * gap of unknown length * 10395 contig of 3827 bp in length * gap of unknown length * 14222 contig of 3522 bp in length * gap of unknown length * 17744 contig of 4825 bp in length * gap of unknown length * 22569 contig of 3725 bp in length * gap of unknown length * 26294 contig of 7302 bp in length * gap of unknown length * 33596 contig of 4984 bp in length * gap of unknown length * 38580 contig of 6259 bp in length * gap of unknown length * 44839 contig of 3662 bp in length * gap of unknown length

\* gap of unknown length  
\* 48501 55257: contig of 6757 bp in length  
\* gap of unknown length  
\* 55258 65262: contig of 10005 bp in length.

## FEATURES

source  
1. .65262  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="F20L2"  
BASE COUNT 21174 a 12169 c 12081 g 19838 t  
ORIGIN

Query Match 71.3%; Score 667.8; DB 2; Length 65262;

Best Local Similarity 82.5%; Pred. No. 5e-186;

Matches 765; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1 ATGCTCACAATTCGCAAAATCTTCGTCGGGTATCGTGGGTGTTGGTGGATCGCC 60

Db 9002 ATGTTAGAGAAATCCGCAAGATCTTTGTCGTGACATAGGGGATTTGGATCTGCC 8943

QY 61 ATTGTCGCGAAGCTTCAGGAACAGGTTTCACCAATCTGTTCTTAAACACACGCGAG 120

Db 8942 ATTGTCGCGAAGCTTCAGGATCAGGTTTCACCAATTTGTCCTTGAACCCATTCAGAG 8883

QY 121 CTTGATCTCACTCGTCAAGCGGATGTTTGAATCTCTTTCTCAAGAGAACCGAGTTTAT 180

Db 8882 CTTGATCTCACTAGCAATCAGATGTTTGAATCTTTCTTCTACAGAGAACCGAGTTTAT 8823

QY 181 GTATCTCAGCAGCAGTAAAGTTGGTGGTATTCACGCTAACACACCTATCTCTGTGAT 240

Db 8822 GTATCTCAGCAGTAAAGTTGGTGGTATTCACGCTAACACACCTATCTCTGTGAT 8763

QY 241 TTCAATGCTCAATCTCCAGATTCAGACCAATGATCCACTGCTCATATGACGCGT 300

Db 8762 TTCAATGCTCAATCTCCAGATTCAGACCAATGATCCACTGCTCATATGACGCGT 8703

QY 301 GTGAAGAAGCTCTCTTCTCTGGATCATCTGATTTACCTTAAATTTGCTCTCAGCCA 360

Db 8702 GTGAAGAAGCTCTCTTCTCTGGATCATCTGATTTACCTTAAATTTGCTCTCAGCCA 8643

QY 361 ATTCCTGAGTCTGTTGTTAAAGCATGCTTGAACCAATGATGATGATGATGATGAT 420

Db 8642 ATTCCTGAGTCTGTTGTTAAAGCATGCTTGAACCAATGATGATGATGATGATGAT 8583

QY 421 GCTAAGATCGCTGGATTAAGATCTGTCAGGCTTATAGGATTCAGACGATCGGATGCA 480

Db 8582 GCTAAGATCGCTGGATTAAGATCTGTCAGGCTTATAGGATTCAGACGATCGGATGCA 8523

QY 481 ATCTCTGGCATGCTTAACTCTATGTCCTTAAATGACAAATTTCCACCGGAGAAATCT 540

Db 8522 ATCTCTGGCATGCTTAACTCTATGTCCTTAAATGACAAATTTCCACCGGAGAAATCT 8463

QY 541 CATGTCCTCTCTCTATGAGAGGTTCCAGCGCGGAAAGTGAATGAGCGGAGAA 600

Db 8462 CATGTCCTCTCTCTATGAGAGGTTCCAGCGCGGAAAGTGAATGAGCGGAGAA 8403

QY 601 GTTGTGCTGGGGTACAGTACTGCTGAGGAGTTCTTGCATCTTGCATGATTTGGCT 660

Db 8402 GTTGTGCTGGGGTACAGTACTGCTGAGGAGTTCTTGCATCTTGCATGATTTGGCT 8343

QY 661 GATGCTTGTGTTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 720

Db 8342 GATGCTTGTGTTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8283

QY 721 GGTCAAGAGTACTATGAGAGTGGCTGAGTGGTGAAGAGGTTGTTGGTTTGA 780

Db 8282 GGTGTTGAAGTACGATCAAGAAATTTGGCTGAGTGGTGAAGAGGTTGTTGGTTTGA 8223

QY 781 GCGAAGCTGGATGGATGCTCAAGCCAGATGGCACACCGAGGAACTTATGGACAGC 840

Db 8222 GCGAAGCTGGATGGATGCTCAAGCCAGATGGCACACCGAGGAACTTATGGATAGC 8163

QY 841 TCAAAGCTCGCTTTGGTTCGACACCTAAAGTTTCTCTAGAGATGCTGAGCAA 900

Db 8162 TCGAAGCTTCGCTTTGGATGACACCGCAAGATTCGCTTAAAGATGCTCTCCAA 8103

QY 901 ACTTATGATGTTTGAAGATGTT 927

Db 8102 ACTTATGATGTTTGAAGATGTT 8076

## RESULT 7

F2H15/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT



```

/codon_start=1
/evidence=not_experimental
/protein_id="AAF97258.1"
/db_xref="GI:9665056"
/translation="MDKSIPIKAIACVACVRPDLVDDVTQNTSTIPTMVSPVLTNMPESA
TSPILMVPRLRTIWPNSKEWYDGDGSPSTGPIKREASDNTNDTAHNTFAPPPEMVVIP
LITIPSDSNYSNCDAGAPSTGVPMVRGRGPKSGKSNSTPTEPKPKYVFNDSLSKVT
SGNDFSEITEATEETGNOEIVDSVMRFDVARRLCQNHHPEDILITTAGNCTKMGV
KNTRRRIAGPIHVGIDIFYYGEMCLVGLHKSNGYGGIDFFTAASAVEGHAAMCVV
TAGQYDEGTEDTLTIYSGQGTGVGNARDQEMKGNLAESVSKGNDVRVRGVI
HPEHNNKIYIVDGMVLSKFWFTVTKSGKFEFRKLVKPKNOPPAYAIWKTVLENLRN
HOLIDSROQFIDLELSFGAELRLVPLNEVEDDDKTIIPEDFDYIPSOCHSGMMWHEFH
FRQSLGQNCQRHQPCHQNCCTVQRNGDLLPYHNNILVCRPLIYECGGSCPCPDHC
PRLVOTGLHLEVFTRKNCGLRSDPDIRAGTPICEFAGRKTKKEVEDDDYLF
DTSKIYQRFNRYVELELLEDEWEQSEFINLPTQVLISAKSGKNGVRPMNHSXPNV
FWQPIEENHFLNSGYLLIGLFAMKHIIPMTTELYDYGVCSVERSEDEGLVCPYLSSS
LWPSSEYIRGLNSKRAWYKIKYKLASQNGVSGLDVSKDEPEKLEQEOGDFGFK
EKLSDVLIDEKLEEVSDCRTATTSRSHTDVSSQTHQTPESRPTPTICDDDTFVS
VSGISRDVSNLIPEATETPASVQEKMANTRFSNNSYKNGODEFFIEDFDVGPMDTI
DYIMTFREDPFDNDLLYAWDRDKTKQRSFKKIMDAIKRKRREKEYEQOLAIWTFG
DADMGCDLVNDKEQSTSIDSKSQTNPNVSVSEDSWEIL"
/complement(6878..8287)
/gene="F2H15.2"
/complement(join(6878..6934,7022..8287))
/gene="F2H15.2"
/notes="Contains similarity to female sterile
homeotic-related protein Frg-1 from Mus musculus
gb|AF045462 and contains a bromodomain PF|00439."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97259.1"
/db_xref="GI:9665057"
/translation="MSSEHISGGGASKTKKHWSQNRKPKMGVSRQSVPLVSPS
NSFASGDDHMLKISLSISKLEENLKLKSELDVRSLLIKRFDPEANPGGSMKAS
GVGRSKYKTKGGGKSGCHGADGTVOIFKNCSLLTKLMKHSAAWVFNVDVAKG
LGLHYDNIYKSPMDLGTVKLGLSKLVKSPLDFAEDVELTFNNALVNPICHDVYRE
AELLNMEDKWSIEMQDNLHRRKPKTRDIEFPAPPSIAPIVEPLPAIVSPSPS
SPPPPPPPVAVLENTWREESMTPIPEPEVAITAPEKAESEEPVNNRDLTLEE
KRLSEELQDLYDKLETVVQIYIKKSNPELSQKDEILDIDSLDINTLWELRYFTVG
YKESLSKNEARHFGFSERDAESVHNSIQBPTTLVSGTTSRVTESGKL"
/notes="Similar to LINE element"
/misc_feature
9839..10402
/gene="F2H15.3"
/join(9839..10049,10191..10402)
/gene="F2H15.3"
/notes="Contains similarity to blue copper-binding protein
III from Arabidopsis thaliana gb|U65650 and contains a
Plastocyanin-like PF|02298 domain."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97260.1"
/db_xref="GI:9665058"
/translation="WAQSGHVSYYAVTVPIAIVMTVLCLFLANAVTVARRPTTVIG
GDGMDDPVPMOTWARGKTFYAGDILEFYDORFNLIIVNRTGYETCEANVGALEYS
SGDDKIQUNYNGFYFTPTPDCTTGLKMAIKALAPR"
/notes="Identical to beta-tonoplast intrinsic protein
(beta-TIP) from Arabidopsis thaliana gb|AF026275 and
contains a MIP (major intrinsic protein) PF|00230 domain.
ESTs gb|R64952, gb|A1999191 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97261.1"
/db_xref="GI:9665059"
/translation="WATSARRAYGFGRADEATHPDPSIRATLAEFLSTFVFPVAGRGSII
LALDKLYMDTAHTGNTNPGVLVALAHALFAAVSAAINVSGSHVNPVPAFTAAII
GGRISVIRAIYVWQAIIACILLIATNGLRPVGFHVASGVSELHGLMEIILT
FALVYVWSTADPRGSGITAPLAIGLIVGANILVGGFFDQASKNPARFAPGALVG
WRWSNHWIYVWGPFGGALALYIYMTIIPSVNEPPHSTHQPAPEDY"

```

```

gene
CDS
complement(12762..16126)
/gene="F2H15.5"
/complement(join(12762..13436,13526..13720,13893..13987,
14151..14499,14602..15232,15319..15560,15920..16126))
/gene="F2H15.5"
/notes="Strong similarity to a hypothetical protein
T18K17.13 gi|659861 from Arabidopsis thaliana BAC T18K17
gb|AC010556 and contains a PH PF|00169 domain."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97262.1"
/db_xref="GI:9665060"
/translation="NFSFVGFVFLSGFGLGILAVLSAAGFWLLKLNLRKRDRIE
SKPSPDPSIKDPNPRESDFCINKQGVVWILDEGLKNMKKEPKPEQKRRKGLLEI
HPLRPALIKHKLILSDADSTSETTVLSIGCSIEAVSGSDLPTRKWAQRPFKYVES
KIPALYKGNQVFIYVLETWSKWCALRLASCENQERFIWYTKLKEDFRNYYVTS
LNVAVSPKPSLGSFETLDKNGTDSKSVRLFLKFKRSNRDRKTVSHHGS
SSCKCPGKNMNTDDTDPFISRSVSHSHISGVSHLSGVSDGSEKVDMDGTLAL
NLLISRLPDLKRTKGVKNSVQARIQRLSNMRTSYIGELICSDVDGTGNLPHIHAT
RVLPMEMSGVWAFELDIYSGDVVIDVETRVDIRVLOQGITDTRLQPRSSGVVPSN
FAEGYDEFEKQLVFFVTNAGEVKNDESRSRGTKAAPNGVSRWKSILKTTAEQVQS
VPISLIRVSSLRGLRVHMKPPSDQLMFGFTSMFDIEFLDASSVGEHKTINSHVAM
FLINREKTAIREAVVLPNCESITIPWMTAEKDDVQKAAAPFWMLNQENDHNTSHATE
ARSKDKPTSPSCLOAQEQMORTANQKIISEIGTLASSSCAQSQVQKAAATAFQKQ
NTEARAIMSTPLSNSTVTIENDKSLEELKTLPLVPSNKNQETNSRQSSREVSAVQS
PSRSVASEEDDSRGKQGRARRMLDGLKMGKLEKRRHMEKRSQIVKVRGSPS"
/complement(17129..19183)
/gene="F2H15.6"
/complement(join(17129..17397,17526..17617,17707..17792,
17894..18128,18214..18431,19070..19183))
/gene="F2H15.6"
/notes="Strong similarity to a hypothetical protein
T18K17.12 gi|6598871 from Arabidopsis thaliana BAC T18K17
gb|AC010556. Est gb|A1996552 comes from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97263.1"
/db_xref="GI:9665061"
/translation="NVGERMWSSTSKLRNLERFLRGITPKPFSLSQSCKNDL
NSLWTHENKDEIYFRLSDLWDFPSPAGLGVKVDLNNESVMQYVYVYSAIQIYS
TNKSTAIIRIHSDVDCESECDSDSEIKLSRSMSSGSKTWDSDGSGVEIDQYS
SLMRKLSIDIQYFESYKPLHVRPLTAKVNLAEKYPLGLSTLRSVDLSPASWLATAM
XPIYHPSRKTDLSTCLFSLYHTLSSAPQGNLIEGDDINETKRETELCFDEGPVTK
SIPLAPFGLVSYKQLQDLWRNQECCDQGRIVYLRSAADSWLKQLNQLVQDHHDSFFSMN
MSL"
/notes="Contains similarity to ATP dependent transmembrane
transporter protein (wh3) from Bombyx mori gb|AF229609 and
contains an ABC transporter PF|00005 domain. ESTs
gb|Z18062, gb|A199375, gb|N96732, gb|F14058, gb|AV528782,
gb|AV559526, gb|AV556190, gb|AV562800, gb|AV559560,
gb|AV523165, gb|AV565094, gb|AV566285 come from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97264.1"
/db_xref="GI:9665062"
/translation="MEIASROQTTVPVSGGNFPVGLSPLSBAIWRKAPTEFVG
DVSARLTWQDLTVMTMGDETQNVLEGLTGYAEPGSLTALMGPSGSGKSTWLDALAS
RLAANAFSGTVLLNGRKTLSFGTAAYVTDQDNLIGTLTVRETTHYSARVRLPDKML"

```

```

Query Match 71.3%; Score 667.8; DB 8; Length 75133;
Best Local Similarity 82.5%; Pred. No. 5.1e-186;
Matches 765; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGTGTTGATGCTGCC 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 36433 ATGTTAGAGAAATCCGCCAAGATCTTTGTTCGTGGACATAGGGGATTGTTGATGCTGCC 36374

```

```
Qy 61 ATTGTCGCGAAGCTTTCAGGAACAGGTTTTCACCAATCTCGTTCTTAAACACACGCGGAG 120
Db 36373 ATTGTCGCGAAGCTTTCAGGAACAGGTTTTCACCAATCTCGTTCTTAAACACACGAG 36314
Qy 121 CTTGATCTCACTCGTCAAGCGGATTTGAATCTCTTTCTCAAGAGAGCGAGTTTAT 180
Db 36313 CTTGATCTCACTAGCAATCAGATGTTGAATCTTTCTTCTCAAGAGAACAGTTTAT 36254
Qy 181 GTAATCTCAGCAGCTAAAGTTGGTGTATTCACGCTAACACACCTATCTCTGCTGAT 240
Db 36253 GTATCTCTAGCTGCTGCTAAAGTCGGTGGATTCATGCTAACACACATACCCAGCTGAT 36194
Qy 241 TTCATCTGCTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCTCATATGACGCGT 300
Db 36193 TTCATCTGCTCAATCTCCAAATCCAGACCAATGTGATCCACTCTGCTTATACACAGGT 36134
Qy 301 GTGAAGAAGCTTCTCTTCTTGATCATCTGATTTACCTTAATACCTTAATGCTCTCAGCCA 360
Db 36133 GTGAAGAAGCTTCTCTTCTTGATCATCTGATTTACCTTAATACCTTAATGCTCTCAGCG 36074
Qy 361 ATTCCTGAGTCTGTTGTTAAACAGATCGCTTGAACCACTAATGAGTGTATGCTATT 420
Db 36073 ATTCCTGAGTCTGCTTCTTACCGTCCGCTCGAACCAGAACCAATGATGATGCGATC 36014
Qy 421 GCTAAGATCTCGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGATGGGATGCA 480
Db 36013 GCCAAGATCGGGTATTAGATGTGTCAAGCTTACCGGCTTACGATCAGTGGGATGT 35954
Qy 481 ATCTGCGGATGCTTACTAATCTTATGTCCTTAATGACAAATTTCCACCGGAGAAATCT 540
Db 35953 ATCTGCGGATGCTTACTAATCTTATGTCCTTAATGACAAATTTCCACCGGAGAAATCT 35894
Qy 541 CATGTGCTTCTCTTATGAGGATTCAGACGAGGCTTCCAGCGGAGGAAATGATGAGCGGAG 600
Db 35893 CATGTGCTTCTCTTATGAGGATTCAGACGAGGCTTCCAGCGGAGGAAATGATGAGCGGAG 35834
Qy 601 GTTGTGCTGCGGATCACAGTATCTGCTGAGGAGTTCTTGATGATGATTTGGCT 660
Db 35833 GTTGTGCTGCGGATCACAGTATCTGCTGAGGAGTTCTTGATGATGATTTGGCT 35774
Qy 661 GATGCTTGTGTTTCTGCTGATTCGATACAGCGGTTGGAGCATTTAAACATTTGGAAT 720
Db 35773 GATGCTTGTGTTTCTGCTGATTCGATACAGCGGTTGGAGCATTTAAACATTTGGAAT 35714
Qy 721 GGTCAAGAGTGAATTAAGAGTTGGCTGAGTTCGTAAGAGGTTCTGCTGTTTGA 780
Db 35713 GGTCAAGAGTGAATTAAGAGTTGGCTGAGTTCGTAAGAGGTTCTGCTGTTTGA 35654
Qy 781 GGAAGCTTGGATGGGATTTGCACTAAGCCAGATGGCACACCGAGGAAACTTATGGACGC 840
Db 35653 GGAAGCTTGGATGGGATTTGCACTAAGCCAGATGGCACACCGAGGAAACTTATGGATAGC 35594
Qy 841 TCAAGCTCGGCTTGTGGTTGGAACCTTAAGTTTCTTATAGATGGTCTGAGCCAA 900
Db 35593 TCAAGCTCGGCTTGTGGTTGGAACCTTAAGTTTCTTATAGATGGTCTGAGCCAA 35534
Qy 901 ACTTATGATGGTATTTGAGATGTT 927
Db 35533 ACTTATGATGGTATTTGAGATGTT 35507
```

```
RESULT 8
AY089031
LOCUS
DEFINITION
Arabidopsis thaliana clone 1353 bp mRNA linear PLN 21-JUN-2002
ACCESSION
AY089031
VERSION
AY089031.1 GI:21407805
KEYWORDS
FLI CDNA.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

```
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1353)
Haas, B.J., Volfovsky, N., Town, C.D., Trukhan, M., Alexandrov, N.,
Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. (2002) In press
2 (bases 1 to 1353)
Brover, V., Trukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1353)
Brover, V., Trukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the Ws or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genset carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
Location/Qualifiers
1.1353
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="19576"
BASE COUNT 371 a 288 c 293 g 401 t
ORIGIN
Query Match 70.4%; Score 659; DB 8; Length 1353;
Best Local Similarity 82.6%; Pred. No. 1.2e-183;
Matches 766; Conservative 0; Mismatches 160; Indels 1; Gaps 1;
Qy 1 ATGCTGCAAAATCGCCAAATCTTCCTCGGGTCAATCGTGGTTCGTTGGATCTGCC 60
Db 199 ATGTTAGAGAAATCGCCAAATCTTTGTTGCTGCGCATAGGGATTCGTTGGATCCGC 258
Qy 61 ATTGTCGCGAAGCTTTCAGGAACAGGTTTTCACCAATCTCGTTCTTAAACACACGCGGAG 120
Db 259 ATAGTCCGCAAGCTTCCAGGATCAGGGTTTCAACCAATCTCGTCTTCGAAACATTCGCGAG 318
Qy 121 CTTGATCTCACTCGTCAAGCGGATTTGAATCTCTTTCTTCAAGAGAGCGAGTTTAT 180
Db 319 CTTGATCTCACTAGCAATCAGATGTTGAATCTTCTTGTCTACAGAGAACAGTTTAT 378
Qy 181 GTAATCTCAGCAGCTAAAGTTGGTGTATTCACGCTAAACACACTATCTCTGCTGAT 240
Db 379 GTCATCTCGCTCGGGCTTAAAGTCGGTGAATTCATGCAACACACATACCCAGCTGAT 438
Qy 241 TTCATCTGCTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCTCATATGACGCGT 300
Db 439 TTCATCTGCTCAATCTCCAAATCCAGACCAATGTGATCCACTCTGCTTATTAAGCAGCGT 498
Qy 301 GTGAAGAAGCTTCTCTTCTTGATCATCTGCAATTTACCCCTAAATTTGCTCTCAGCCA 360
Db 499 GTGAAGAAGCTTCTCTTCTTGATCATCTTGGTTTCACTTGGCATACCCCAATTTGCTCTCAGCG 558
Qy 361 ATTCCTGAGTCTGTTGTTAAACAGATCGGTTGAAACCAACTAATGAGTGTATGCTATT 420
Db 559 ATTCCTGAGTCTGCTTCTTACCGTCCGCTCGAACCAGAACCAATGATGATGCGATC 618
```





Db 23227 CCAAGATACAGGAGATGGGTGAAGCCCAAGGTCCCTCAAGGAGGGGCTCGTGGAGA 23286  
QY 902 CTTATGATTGGTATTGGAAGAATGT 926  
Db 23287 CATACAAGTGGTATGTCGAGAATGT 23311

RESULT 10  
AX172657  
LOCUS AX172657 1490 bp DNA linear PAT 03-JUL-2001  
DEFINITION Sequence 147 from Patent WO0144476.  
ACCESSION AX172657  
VERSION AX172657.1 GI:14597769  
KEYWORDS  
SOURCE  
ORGANISM  
Phycomitrella patens.  
Phycomitrella patens.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
REFERENCE  
AUTHORS Lerchl, J., Renz, A., Ehrhardt, T., Reindl, A., Cirpus, P., Bischoff, P.,  
Frank, M., Freund, A., Duwenig, E., Schmidt, R. M. and Reski, R.  
TITLE Moss genes from phycomitrella patens encoding proteins involved in  
the synthesis of carbohydrates  
JOURNAL Patent: WO 0144476-A 147 21-JUN-2001;  
BASF Plant Science GmbH (DE)  
FEATURES  
source  
1. 1490  
/organism="Phycomitrella patens"  
/db\_xref="taxon:3218"  
347..1279  
/notes="c\_pp001064043r"  
/codon\_start=1  
/protein\_id="CAC43708.1"  
/db\_xref="GI:14597770"  
/translation="MGVDDKDAKIFVAGHRLVGAAVVRALKDGVNNLVKMKTHKBLDL  
TRQHEFFDTEKPAVYILAAKVGVIHANSTYPAEFIAVNLQIOTNVDIAAYKGVKK  
LFLGSSCIYKPAQVPIVEESLLTGPLTEATNEWAKIAGIKMCOAYRLOYNFDAI  
SGMPTNLGYPHDNFHPENSHVLPALIRRPHEAKVNGAKEVVMVVGSGSPREPLHVDLL  
ARATVLLONYSAHEHVNMGSGSEVSIKELAEKVEVVGFOQLTWDISKPDGTPRKL  
IDSSKLANGWQARIPLEKGLAETKYWCYNV"  
BASE COUNT 369 a 308 c 410 g 402 t 1 others  
ORIGIN

Query Match 41.8%; Score 391; DB 6; Length 1490;  
Best Local Similarity 65.0%; Pred. No. 2.6e-104;  
Matches 596; Conservative 0; Mismatches 315; Indels 6; Gaps 1;

QY 7 GACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTTGGTTGGATCTGCCATTGTC 66  
Db 356 GACAAGGAGCCCAAGATCTTTGTTGCTGGACACCGAGGTCTAGTAGTGGCGCTGTGTT 415  
QY 67 CGCAAGCTTCAGGAACAAGTTTCCACCATCTCGTTCTTTAAACACACCGCCGAGCTTGT 126  
Db 416 CGTGTCTTGAAGAAGATGGTTATACAAATTTGGTGATGAAGACTCATAAAGAAGCTAGAT 475  
QY 127 CTCACTCGTCAGCCCATGTTTGAATCTTCTTTCTTCTCAGAGAAGCCAGCTTTATGTAATC 186  
Db 476 CTTACCGTTCAGCACAGS-----AATTTTCGACACGGAAGAACACGTCAGTCATC 529  
QY 187 CTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAACCAACCATATCTCTGCTGATTTCAAT 246  
Db 530 CTAGCAGCTCGAAGGTGGGAGGCATTACGCAACACAGTACTTACCTCGACAGTTTCATT 589  
QY 247 GGTGTCAATCTCCAGATTTCAGACCATGATCCACTCTGCATATGACGACCGGTGGAAG 306  
Db 590 GCGTGAATCTGCAGATCAAAACGAATGTCATCGATGCTGCTGCTTACAAGTCTCGGTGAAG 649  
QY 307 AAGCTTCTCTCTCTTGGATCATCTCTGATTTACCTTAAATTTGCTCTCTCAGCCAATTCT 366  
Db 650 AAGCTCTTCTTCTGGCTCTCTGGTGATCTACCCCAAGTTTGGCCAGGTACCAATCGTT 709  
QY 367 GAGTCTGCTTTGTTAAACAGCATCGCTTGAAACCAACTAATGATGGTGTATGCTATTGCTAAG 426

Db 710 GAGGAGTCTCTCTCACAGGGCTTTTGAAGCTACAAACGAGTGTGTCTGTAGCAAG 769  
QY 427 ATCCGCTGGGATTAAGACTTTGTCAGGCTTATAGGATTCAGCAGGATGGGATCAATCTCT 486  
Db 770 ATTCAGGAATCAAAATGTGCCAGGCTTACAGGCTGCAGATTAATTTTCGACCCCAATTC 829  
QY 487 GGCATGCCCTACTAATCTCTATGGTCTTAATGACAATTTCCACCCGGAGAAATTTCTCATGTG 546  
Db 830 GGAATGCCGACAAACCTCTACGGTCCCCACGACAAATTTCCATCCCGAGNAATCCCACTG 889  
QY 547 CTTCTCTGCTTTATGAGGAGTTTCAAGGAGGAAAGTGAATGAGCGGAGGAGGAAGTTGTG 606  
Db 890 TTGCCAGGCTTGTATCAGACGCTTTTCAAGAGGCTAAGTGAAGCGGCTAAGGAAGTGTG 949  
QY 607 GTGTGGGTACAGGTAGTCCGTGAGGAGTCTTTCGATGTTGATGATTGGCTGATGCT 666  
Db 950 GTGTGGGATCAGGTTCCCATTTCCGTGAGTTTCTTCAGTGAGGAGCTTTGGCAGAGGCA 1009  
QY 667 TGTGTTTTCTTGTCTGATCGATACAGCGGTTGGAGCATGTTAAACATTGGAAAGTGTCAA 726  
Db 1010 ACAGTATTCTTGTCTGACAAATTTACTCCGCGCATGAGCATGTCAACATGGGCACTGCTCT 1069  
QY 727 GAAGTGACTATTAGAGAGTTGGCTGAGTTGTTGTTGAAAGAGTTGTTGTTTGAAGGGAAG 786  
Db 1070 GAGGTCTCAATCAGGAAGTCCGCGAAATGGTGAAGGAAGTGGTTGGATTTTCAGGGGCG 1129  
QY 787 CTTGATGGGATGTCATTAAGCCAGATGGCACACCGGAGGAACTTTATGGACAGCTCAAG 846  
Db 1130 CTGACATGGGATCTTTTAAGCCTGATGGAATCCACGAAAGCTCATCGATAGGAGCAAA 1189  
QY 847 CTGCGGTCTTTGGGTTGGACACCTAAGGTCTTCTTTAGAGATGGTCTCAGCCCAACTTAT 906  
Db 1190 CTTGCCAATGCGGTGGCAAGCGAATTTCCCTCAAGGAAGGATTGGCAGAGACTTAC 1249  
QY 907 GATTGGTATTGGAAGAA 923  
Db 1250 AATGGTACTGTGAGAA 1266

RESULT 11  
AE013290/c  
LOCUS AE013290 10029 bp DNA linear BCF 17-MAY-2002  
DEFINITION Methanosarcina mazei strain Goel, section 72 of 379 of the complete  
genome.  
ACCESSION AE013290 AE008384  
VERSION AE013290.1 GI:20905046  
KEYWORDS  
SOURCE  
ORGANISM  
Methanosarcina mazei Goel.  
Methanosarcina mazei Goel  
Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
Methanosarcinaceae; Methanosarcina.  
REFERENCE  
AUTHORS Deppe, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A.,  
Martinez-Arias, R., Henne, A., Wier, A., Baeumer, S., Jacob, C.,  
Brueggemann, H., Lienard, T., Christmann, A., Boemcke, M., Steckel, S.,  
Bhattacharya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,  
Gunsalus, R. P., Fritz, H.-J. and Gottschalk, G.  
TITLE The genome of Methanosarcina mazei: Evidence for lateral gene  
transfer between Bacteria and Archaea  
J. Mol. Microbiol. Biotechnol. (2002) In press  
REFERENCE  
AUTHORS Deppe, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A.,  
Martinez-Arias, R., Henne, A., Wier, A., Baeumer, S., Jacob, C.,  
Brueggemann, H., Lienard, T., Christmann, A., Boemcke, M., Steckel, S.,  
Bhattacharya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,  
Gunsalus, R. P., Fritz, H.-J. and Gottschalk, G.  
TITLE Direct Submission  
JOURNAL Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute  
for Microbiology and Genetics, Grisebachstrasse 8, Goettingen  
37077, Germany  
FEATURES  
Location/Qualifiers  
1. 10029  
source



```

GUDMTPEMVEKQANARKYGSNVFROQDIEALPVEDSSVDVIIISNCVINLAPDKEK
VFREAPRVKLGRLVYISDLVLDLDELPEDLKNDRLDLAGCAGVAGLKEVYKLLKRAAG
FSVEILAEPDIDSKSQYKGLPVESLKLAWV"
complement (9368. .9787)
/gene="MM0662"
complement (9368. .9787)
/gene="MM0662"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Arsr family"
/protein_id="AAM30358.1"
/db_xref="GI:20905056"
/translation="MLYNKVFKEGDELTRILSYLLEHSYCADFSLSLKQDOTT
VSKHLVLEADILKYKGRNIIYIKNRETECLRLSLGIRDKVPCCDRQQAASMLCF
SVVAENRAGNLIDEEIEDRAESEAKDKSVNLQAKK"

BASE COUNT 3063 a 2058 c 1687 g 3221 t
ORIGIN
Query Match 37.8%; Score 354; DB 1; Length 10029;
Best Local Similarity 62.5%; Pred. No. 3e-93;
Matches 574; Conservative 0; Mismatches 335; Indels 9; Gaps 1;

QY 7 GACAAATCTCCAAAATCTTCGTCGGCGGTGTCATCGTGGTGTGTTGGATCGCATGTC 66
Db 5723 GATAAAGGTCAAAAATTTATGTGGTGGCCACCGAGGCTTGTAGGTTGAGCTCTGAAA 5664

QY 67 CGCAAGCTTCAGGAACAAGGTTTCAACAATCTCGTCTTAAACACACCGCGAGCTTGTAT 126
Db 5663 AGAAACTTGAGTCAAAAGATTAATCAATCTTATCTTCGAACTCATCGAGAGCTTGTAT 5604

QY 127 CTCACCTCGTCAAGCCGATGTTGAATCCTCTTCTCAAGAGAGCCAGGTTTATGTAATC 186
Db 5603 CTCACCTAATCAGCAGCTGTTATGAATTTTGAACGAGAAAGCGGAGTATGTTT 5544

QY 187 CTAGCAGCAGCTAAAGTTGGTGTATTCACGCTAAACACCTATCTCTGCTGATTTTCATT 246
Db 5543 TTAGCGCTGCAAAAGTTGGTGAATTTCTTGCCTAATACTATCTCTGCGAGTTTCATT 5484

QY 247 GGTGTCAATCTCAGATTCAGACCAATGTATCCACTTCGATATGATGACGAGCGTGAAG 306
Db 5483 TAGCAAAATCTTATGATCGAAGCAAAATATTTACGCTTCCTACAGTGGCGGTAATA 5424

QY 307 AAGCTTCTCTCTCTGATCATCTCGATTTACCTTAAATTTGCTCTCAGCCAAATTCCT 366
Db 5423 AACTACTTTTCTGGTCTCTCTGATTTATCTTATCTTAACTCGCTCCCACTCTCAAA 5364

QY 367 GAGTCTGCTTTGTAACAGCATCGCTTGAACCACTAATGAGTGGTATGCTATGCTAAG 426
Db 5363 GAAGAATATCTATTTGACCGGCCCTCTTGAAGAAACAATGAAGCATATGCGAGTTGCAAG 5304

QY 427 ATCGCTGGGATTAAGACTTCTCAGGCTTATAGGATTCAGCAGGATGGGATGCAATCTCT 486
Db 5303 ATCGAGGAATAAGCTTTGCAACACTAATACGACGATGAGGACTAATCTTCAATATCG 5244

QY 487 GGCATGCTACTAATCTCTATGTCCTTAATGACAAATTTCCACCGGAGAAATTTCAATGTG 546
Db 5243 GTAATGCGACCAATCTTTATGCGCTTAATGATAATTTGACCTTGAGACTTCACATGA 5184

QY 547 CTTCTGCTCTTATAGGAGGTTCCAGGAGCGGAAAGTGAATGAGCGGAGGAGTTGTG 606
Db 5183 ATGCTGTCATGTTGAAGGAGTTCCATGAAGCAAGCTGAATTAATGAGCGCTGAAGTTGA 5124

QY 607 GTCTGGGGTACAGTATGCTGAGGAGTTCTTCGATGTTGATGATGATTTGGCTGATGCT 666
Db 5123 ATCTGGGGTACCGGTAAACCATACAGGAGTTTCTTACGATGATGATATGCTGACGCC 5064

QY 667 TGTGTTTCTTCTGCTG-----ATCGATACAGCGGGTTGGAGCATGTTAAACATGGA 717
Db 5063 TGTGTGTATTATGGAATACTTTAATACGATGACATGAGGAAATTTGTCAATATCGG 5004

QY 718 AGTGGTCAGAAGTGTATAGAGAGTTGGCTGAGTTGGTGTGAAGAGTTGTTGTTT 777
Db 5003 GTAGGGAAGATATTACAATAGTGAATTAATAGCTGAATTAATCAAGAGATCGTAGGTTT 4944

```

```

778 GAAGGAGAGCTTGGATGGATGCTAAAGCCAGATGGCACACCGAGGAAACTTATGGAC 837
4943 AAGGTGAATAGAAAGATCTTTCNAAGCTGACGGTACTCCACAGAGCTGCTTAT 4884

838 AGCTCAAAAGCTCGCTCTTTGGTTGGACACCTAAAGTTTCTCTTAGAGATGCTCTGAGC 897
4883 ATTACAAAATTCAGCTCACTTGGTGGAAAGCTAATATCTCATTAAGAGATGAATAGA 4824

898 CAAACTTATGATGCTAT 915
4823 CAACTTATGAATGAT 4806

RESULT 12
AE007718/c
LOCUS
DEFINITION
  Clostridium acetobutylicum ATCC824 section 206 of 356 of the
  complete genome.
ACCESSION
  AE007718 AE001437
VERSION
  AE007718.1 GI:15025171
KEYWORDS
  Clostridium acetobutylicum.
SOURCE
  Clostridium acetobutylicum
  Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
  Clostridium.
REFERENCE
  1 (bases 1 to 8458)
  Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
  Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
  Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
  Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  Genome sequence and comparative analysis of the solvent-producing
  bacterium Clostridium acetobutylicum
  J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
MEDLINE
  11466286
REFERENCE
  2 (bases 1 to 8458)
  Childress,D., Zeng,Q. and Smith,D.R.
  Direct Submission
  Submitted (24-JUL-2001) GTC Sequencing Center Production,
  Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
  Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
  Location/Qualifiers
  1..8458
   /organism="Clostridium acetobutylicum"
   /strain="ATCC 824"
   /db_xref="ATCC:824"
   /db_xref="taxon:1488"
   complement (157. .1527)
   /gene="CAC2177"
   complement (157. .1527)
   /gene="CAC2177"
   /codon_start=1
   /transl_table=11
   /product="Ucharacterized protein, CGBB homolog"
   /protein_id="AAK80135.1"
   /db_xref="GI:15025172"
   /translation="MLKVLITCLKYDNDKSRGYSFEYENFYKTLIKMDGIELLPDI
   CDGDKKRENNNNLKIIEKRPDILNIDYEDQIKETFLYIKNNITKILVNWFC
   DDQRFESTIKWCDFYCVTYKKAIVKYKELGYENIIFQWACNQYNYPKRDI PF
   KYDVSFVQPHSNRRIINLKQKIEIVACFGYGNEDPNRSRIISQSMIDVFNSSK
   INLNSHLDAPQOIKGRNFPACGAFILTSDEGLSHYIEIGKEVTVVYSSDDM
   VDKIKYLINEKRTIANAGVIRTIKEHTYENLRNLFKVLKDGKDNKQMDLFFY
   RPNYKADLVSLFQNAVGNIIGSGDHTNLIKYKLLIGDIKENTYVFDNSL
   KNGTEYLGGIITHSPEIDELNLDRIIISSEYEEDI FYKLINLITSGINIVKIYNGDKK
   ENLFTD"
   complement (1598. .2752)
   /gene="CAC2178"
   complement (1598. .2752)
   /gene="CAC2178"
   /codon_start=1
   /transl_table=11
   /product="Predicted"

```

```

CDP-4-keto-5-deoxy-D-glucose-3-dehydrase"
/protein_id="AAK80136.1"
/db_xref="GI:15025173"
/translation="MKWKLQENALVKEETESLAETKTADRLTQGSVKVEFAEWSKV
QCKYSVFVNSGSSANLVLLSAVKDYVNDLNTSEVIVPAVTVWTVNISPILOGLKAPV
VDVNLKEDSFDEELKKITDKTKAIFTHLLGFPAWMEKIKNVIKEXDIVLEDCCE
SHGAKYKIKVGNFGLGTFSTFYWGHMTTIEGGMISTNNEEFYKLLCLKRSHGLARE
LPIKHEYKIKYKIDIDFKFLDTGDFNVRSTENAFGLRQIKNLDAPFIEIRDKNYN
MFDFCTKYDEKLTITIKVEGTSLSALPIFKRKLREIVKREKLNELGVETREIISGNL
LQPFLLKYYHNEFEKNANLIHDNGFYIGNNQFVDEIRLKLKLESLDKIM"
complement (2768..3697)
/gene="CAC2179"
complement (2768..3697)
/gene="CAC2179"
/codon_start=1
/transl_table=11
/product="Nucleoside-diphosphate-sugar epimerase"
/protein_id="AAK80137.1"
/db_xref="GI:15025174"
/translation="MKEDSKIYIAGHTGPGVSGAILRLNERRGYKNVVRTHKELDLMH
QBSVKFLFEERKPDYVLSAAKVGIGIQANISNPVDFLMDNLIIEYVNIKNSEVGIEN
LFLAGSSCIYKPEAPQPLKEEYLLSGYLEPTNEGVAIAKISGLKMCVEYSKQYGLNYI
SAMPNLSYMRONFDLKTSHVMAALIRPHEAKVSGSOBISIWGSGEOVREFTYEDL
ADGIIIFLMEHGEKVGELNIGCGDKIKLDLAYKIKDVVGFKGNIIFDKSKPDGMFRK
VNDVSKINSLGWHYKVELDEGKITRYRYLNNC"
complement (3712..4779)
/gene="CAC2180"
complement (3712..4779)
/gene="CAC2180"
/codon_start=1
/transl_table=11
/product="GDP-D-mannose dehydratase"
/protein_id="AAK80138.1"
/db_xref="GI:15025175"
/translation="WKKAIIITGITGQDGSYLTFELLLKGYEVHGIIRASSFNTKRID
YLPEPKTGNKLLFLHHGLDTSNINLLEKVOQTEIYNLAQSHVFSRVPETTV
ETDAVGLFLDAIKELGDKCKFYQASTSELFEGMFGTAPQSEKTFPFRSYPAAALK
YAYMTIVNRYEAGTIFACNGLVFNHESPRRGTFVTRKITRAIANIIEGNODKVLGN
LNAKDQVYAGVYVAMLMQLQDTPKDYVIATGETHTVREVEKAFKAFAGIKIKWIG
QGIIEBKIDLTKGTLVDVNPFRPTVEVLLGLDPSMAERELGWRKRVNPELVVM
VREDLKKTIRITNVDVRELLT"
complement (5047..5334)
/gene="CAC2181"
complement (5047..5334)
/gene="CAC2181"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK80139.1"
/db_xref="GI:15025176"
/translation="WMRLCRLYITVTILENKILELEKIEDRPEKLIROKKOILYTDRF
FNVLSESDREYACNVDTYNLLKSTWVPVFKSKINMLNGLGQYKQKQK"
complement (5386..5502)
/gene="CAC2182"
complement (5386..5502)
/gene="CAC2182"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAK80140.1"
/db_xref="GI:15025177"
/translation="WNKIFLLVCMCIMVKAIVKVCYRINYQANISELSFYIR"
complement (5827..6000)
/gene="CAC2183"
complement (5827..6000)
/gene="CAC2183"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein, possible homoloh of YJFB
B. subtilis"
/protein_id="AAK80141.1"
/db_xref="GI:15025178"
/translation="WDIALLATINSNEKVNEANLAVFQMAVNNCKQNGENTADMIEA

```

```

MDPNRGNNDIRV"
6136..7053
/gene="CAC2184"
6136..7053
/gene="CAC2184"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein, homolog H11244 from
Haemophilus influenzae"
/protein_id="AAK80142.1"
/db_xref="GI:15025179"
/translation="MYKCLIWGVDSYTMAYDKLFEILKKNLSIEALISKDKYAEYI
DGKVIDKTEISNYOFDIILIFNKERYSDIKNEALELGIPEKILNGKVFVTSNDFDK
RYCKLIENPITIIISNNCGGKISNHLGKFNFPINLLELDYMKFLNMDYLLDQE
LYQVDEGDVYSCDIPKSGISGDNKIIIFNHNASFEAKNDWERKKRINKNLFIK
MTLPANNEVKRFDNLVYKNKVCYFPRPMKYKSVFPCRYIWKCKQMARKTSNYDLV
YFVMDTNLQKSDILKMLCGEDDFIREK"
7463..8380
/gene="CAC2185"
7463..8380
/gene="CAC2185"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein, homolog H11244 from
Haemophilus influenzae"
/protein_id="AAK80143.1"
/db_xref="GI:15025180"
/translation="MYKCLIWGVNDEYTLAYDKLFEISKGNLSIEALISKDKYAKYI
DGKEVIDKTEISNYEFDYIILFNKERYSDIKNEALELGIPEKILNGKVFVTSNDFDK
RYCKLIENPITIIISDDCGGLVSYLGPKNFSPFINFYIHNDYIKFLENMDYLLDQE
LKVEQEGNVISCTMPKSLGTGDNKIIIFNHNQASFAEAKNDWERKKRINKNLFIK
MLIKDNEKLVKRFNLPYKNKVCYFPRPMKYKSVAFPRYIWRICINYYAARTSNLSLE
QYTMDSWLEKSCDILKMLCGEDDFIREK"
BASE COUNT 2863 a 1429 c 939 g 3227 t
ORIGIN
Query Match 32.2%; Score 301.8; DB 1; Length 8458;
Best Local Similarity 58.7%; Pred. No. 8.1e-78;
Matches 541; Conservative 0; Mismatches 377; Indels 3; Gaps 1;
QY 11 AATCTGCCAAATCTTCGTCGGGTCATCGTGGTTTGGTTCGATCTGCCATGTCCGCA 70
Db 3690 AAGATAGCAAAATATATATAGCAGGTTCATCTGGTTTGGTTCGTCATATTAGAA 3631
QY 71 AGCTTCAGGAACAAGTTTCCCAATCTCGTTCTTAAACAACACGCCGAGCTTGATCTCA 130
Db 3630 ATTTAGCGGTAGAGATATAAAATGTTGTAGTGAGAACTCATAAAGAACTTGATTGA 3571
QY 131 CTCGTCAGCCGATGTTGAATCCTTCTTTCTCAAGAGAAAGCGATTAATGTAATCCTAG 190
Db 3570 TGCACCAAGAAGTGTAAAGAGTTCCTTGAAGAAGAAAGCCTGATTATGTAGTATTAT 3511
QY 191 CAGCAGCTAAAGTTGGTGGTATTACGCTAAACAACACCTATCTCTGATTTTCATTTGG 250
Db 3510 CAGCTGCAAAAGGTTGGTGAATACAAAGCGAATATTTCGAATCCTGTGGAATTTCT 3451
QY 251 TCAATCTCCAGATTCCAGACCAATGTGACCTCTGTCATATGAGCACCGGTGTGAAGAAG 310
Db 3450 ATAACTTAATAGAAATATAATGATAAAAAATCTTTGAGGTGAGAAATGAAAAAT 3391
QY 311 TTCTCTTCTCTGGATCATCTCGATTTACCCCTAAATTTGCTCTCTCAGCCAAATTCCTGAG 370
Db 3390 TACTATTTTGGAGGAGTTCTTGATATATCTTAAAGAACCAACACAGCCTTTGAAAGAAG 3331
QY 371 CTGCTTTGTTAAACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATTGCTAAGATCG 430
Db 3330 AATATCTGCTTTCAGGATATTTGGAGCCCACTAATGAAGGTTATGCTATAGCTAAGATAT 3271
QY 431 CTGGGATTAAAGACTTGTGAGGCTTATAGGATTACGACCGGATGGGATGCAATCTCTGGCA 490
Db 3270 CAGGCTAAAGATGTGCGAATATATTATAGACAGTATGGAATTAATTACATAAGTGCTA 3211
QY 491 TGCCTACTAATCTCTATGGTCTCTAAATTTCCACCCGAGAAATTTCTCATGTGCTTC 550

```

Db 3210 TGCCTTCAAAATTTGATGGTATGAGAGATAAATTTTGATTTGAAAACTTCTCATGTTATGG 3151  
Qy 551 CTGCTCTTATGAGAGGTTCCAGAGCGGAAGTGAATGAGCGGAGGAGTTGTTGGTGT 610  
Db 3150 CGGCATTAATAAGAGGTTTCATAGGCAAGGTTCCGGGTCTCAGGAATAATCCATAT 3091  
Qy 611 GGGGTACAGGTAGTCCGTTGAGGAGTTCCTTGATGTTGATGATTTGGCTGATGCTTGTG 670  
Db 3090 GGGGAAGTGAGAACAGTATAGAGAAATTTACATATATAGAGGATTTGGCTGATGGAATTA 3031  
Qy 671 TTTTCTTGCTGG---ATCGATACAGCGGTTGAGCAATGTTAACTTGGAAAGTGGTCAAG 727  
Db 3030 TTTTCTTGATGGAACATGGCGAAAAGTGAAGGTTTTTTAAATATAGGTTGTGGCAAG 2971  
Qy 728 AAGTGACTATTAGAGGTTGGCTGAGTTGGTGAAGAGGTTGTTGTTTGAAGGAGC 787  
Db 2970 ATATTAAATAAAGATTTAGCTTTATAAAATAAAGATGTTAGTTGGAATTTAAAGGAATA 2911  
Qy 788 TTGGATGGGATTCACATAGCCAGATGGCACACCGAGGAACTTTATGGACAGCTCAAAGC 847  
Db 2910 TAATTTTGTATAGAGTAAAGCAGATGGAATGTTAGAAAGTTATGGATGTTTCTAAA 2851  
Qy 848 TCGGCTTTGGGTGACACCTAAGGTTTCTCTTAGAGATGTTCTGAGCCAACTTATG 907  
Db 2850 TTAATAGTTTAGGATGTCATTATAAAGTTGAACTTGATGAAGGAATTAACAAAACATATA 2791  
Qy 908 ATTGGTATTGAAGAATGTTT 928  
Db 2790 GGTGGTATTAAATAATGTT 2770

## RESULT 13

LOCUS A63794 1020 bp DNA linear PAT 12-MAR-1998  
DEFINITION Sequence 15 from Patent WO9723624.  
ACCESSION A63794  
VERSION A63794.1 GI:3717366  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1020)  
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,  
Sumar,N. and Ford,J.  
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA  
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY  
JOURNAL Patent: WO 9723624-A 15 03-JUL-1997;  
ST GEORGE S HOSPITAL MEDICAL S (GB)  
COMMENT Other publication AU 1202797 19970717.  
FEATURES  
Location/Qualifiers  
Source 1..1020  
/organism="unidentified"  
/db\_xref="taxon:32644"  
CDS 1..1020  
/note="unnamed protein product; protein sequence is in  
conflict with the conceptual translation"  
/codon start=1  
/protein\_id="CAA03745.1"  
/db\_xref="GI:3717367"  
translation="VRWHTMDHADVAVLGRSKLTPTTGPGLDRATPVYIAGHRLVGS  
ALVRFPEAGFTNLI VRSRDEIDLTDRAATFDVSETRPQVIIDAAAVVGIMANNY  
PADFASENIRIOTNLIDAAVAVRPRLLFLGSCIPKYAPQPIHESALLTGPLEPTN  
DAYATAKIGILQVQVRQYGLAWISAMPTNLYGPDNFSPSGSHLLPALIRRYEEA  
KAGAEVNTMGTPRRELLHVDOLASACIFLEHFDGPNVHVGTVGDHISISIA  
MYATAVCYIGETRWDPDKEDGTPRKLDDVSLRELGRWRPRIALKDIDATVSWYRTNA  
DAVR"

## BASE COUNT

207 a 297 c 305 g 211 t

## ORIGIN

Query Match 31.7%; Score 297; DB 6; Length 1020;

Best Local Similarity 58.2%; Pred. No. 1.7e-76;

Matches 522; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

Qy 29 TCGCGGTCATCGTGGTGTGGATCTGCCAATGTCGCAAGCTTTCAGGAACAAGGTT 88  
Db 104 TCGCGGTCATCGGCGGCTGGTTCGGCTCAGCGCTGTAAGTAGATTTGAGCGCGGGGT 163  
Qy 89 TCACCAATCTCGTTCCTTAAACACACGCGAGCTTGATCTCACTCGTCAAGCCGATGTTG 148  
Db 164 TCACCAATCTCATGTTGCCATCAGCGGATGAGATTGATCTGACGGACCGCGCAACGT 223  
Qy 149 AATCCTCTTTTCTCAAGAGAAGCCAGTTTATGTAATCTAGCAGCAGCTAAAGTTGGTG 208  
Db 224 TTGATTTGCTGTGAGACAAGACCAAGGTGATCATCGATCGCGCGCAGCGGTGGCG 283  
Qy 209 GTATTCACTCAACACCTATCTCTGCTGATTTCATTGCTGTGTCATCTCCAGATTCAGA 268  
Db 284 GCATCATGGCGAATACACCTATCCGCGGACTCTCTTGTCCGAAAACCTCCGNATCCAGA 343  
Qy 269 CCAATGTGATCCACTCTGCATATGAGCAGCGGTGTGAAGAGCTTCTCTTCTTGGATCAT 328  
Db 344 CCAATTTGCTCGACGAGCTGTCGCGTGGTGTGCGCGGCTCCTTTTCTCGGTTCGT 403  
Qy 329 CTTGCATTTACCTTAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCAT 388  
Db 404 CATGCATCTACCGCAAGTACGCTCCGCAACCTATCCAGAGAGTGTCTTTATTGACTGGCC 463  
Qy 389 CGCTTGAACCAACTAATGAGTGTATGCTATTGCTAAGATCGCTGGGATTAAGACTTGTG 448  
Db 464 CTTTGGAGCCCAACCAAGCGGTATGCGATCGCCAGATCGCGGTATCTTGCAGAGTTC 523  
Qy 449 AGGCTTATAGGATCAGCACGAGTGGATGCAATCTTGGCATGCTACTAATCTCTATG 508  
Db 524 AGGCGGTTAGCGGCCAATAATGGCTGGCGTGGATCTCTGCGATGCGGACTTAACCTCTACG 583  
Qy 509 GTCTTAATGACAAATTTCCACCGGAGAAATTCATGTCGCTTCTCTTATGAGAGGT 568  
Db 584 GACCGCGGCAACAATCTCTCCGCTCCGGTTCGCTCTCTTCCGCGGCTCATCCGTCGAT 643  
Qy 569 TCCACGAGCGCAAGTGAATGGAGCGGAGAAATTTGTGTGGGTGAGTACAGGTAGTCCGT 628  
Db 644 ATGAGGAGCCAAAGCTGGTGTGAGAGAGTGCAGNAATTTGGGGACCGGTACTCCGC 703  
Qy 629 TGAGGAGTCTTGTGATGTTGATGTTGGTGTGATGCTGTGCTTTTCTTGTGGATCGAT 688  
Db 704 GCGCGCAACTCTGCTATGTCAGCATCTGGCAGCGCATGCTGTCTCTTTTGAACAT 763  
Qy 689 ACAGCGGTTGGAGCATGTTAATCATTTGAGTGGTCAAGAGTACTATTAGAGTTCG 748  
Db 764 TCGATGGTCCGAAACCAACGTCACGTCGCGGCAACGCGGTGATCAGCATTTAGCAGATCG 823  
Qy 749 CTGAGTTGGTGAAGAGGTTGTTGTTTGAAGGGAAGCTTGGATGGGATTCACCTAAGC 808  
Db 824 CAGACATGTCGCTACGCGGTGGGCTACATCGCGGAACACGTTGGGATCCCACTAAC 883  
Qy 809 CAGATGGCACACCGAGGAACTTATGGAAGCTCAAGCTCGCGTCTTTGGGTGGACAC 868  
Db 884 CCGATGGAACCCCGCGCAAACTATTGGAGCTCTCCGCGCTACGCGAGTTGGGTGGCGCC 943  
Qy 869 CTAAGTTTCTTTAGAGATGCTCTGAGCCAACTTATGATGGTATTGTAAGATG 925  
Db 944 CGCGAATCGCACTGAAGACCGCATCGATGCAACGGTGTGTTGTTACCGCACAAATG 1000

## RESULT 14

MAP223833  
LOCUS MAP223833 4435 bp DNA linear BCT 07-JAN-1999  
DEFINITION Mycobacterium avium paratuberculosis gs[ba,bb,c,d] genes.  
ACCESSION AJ223833  
VERSION AJ223833.1 GI:3550477  
KEYWORDS gsa gene; gsb gene; gsc gene; gsd gene.  
SOURCE Mycobacterium avium subsp. paratuberculosis.  
ORGANISM Mycobacterium avium subsp. paratuberculosis.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium





Db	1815	ATGAGGAAGCCAAAGCTGGTGGTGCAGAGAGGTGACGAATTCGGGGACCGGTACTCCGC	1874
Qy	629	TGAGGGAGTCTTTGTCATGTTGATGATTTGGCTGATGCTTGTGTTTCTTCTGCTCGATCGAT	688
Db	1875	GGCGCAACTTCTGCATGTGCAGGATCTGGCAGCGCATGCCCTGTCTCTTTTGGAAACATT	1934
Qy	689	ACAGCGGGTTGGAGCATGTTTAAACATTGGAAAGTCGTCAAGAAGTGACTATTATAGAGAGTTGG	748
Db	1935	TCGATGTCGCAACCAAGCTCAAGCTGGGCACCGGCGTCGNATCACAGCATTTAGCGAGATCG	1994
Qy	749	CTGAGTTGGTGAAAGAGTTGTTGGTTTTGAAAGGAAGCTTTGGATGGGATTTGCATTAAGC	808
Db	1995	CAGACATGTTGTCGATACGGCGGTGGGGCTACATCGCGCAAAACACGCTTGGGATCCAACTAAAC	2054
Qy	809	CAGATGCCACACCGAGGAAACTTATGACAGCTCAAAAGCTCGCGCTCTTTGGGTGTGGACAC	868
Db	2055	CCGATGGAACCCGCGCAACTATTGACGTCTCCGCGCTACCGGAGTTGGTGTGGCGCC	2114
Qy	869	CTAAGGTTTCTCTTAGAGATGGTCTCGAGCCAAACTTATGATGGTATTTTGAAGAATG	925
Db	2115	CGCGAATCGCACTCAAAAGACGGATCGATGTCACCGGTGCTGGTACCGCACAAATG	2171

Search completed: June 3, 2003, 03:32:04  
Job time : 2641 secs

Qy	869	CTAAGGTTTCTCTTAGAGATGGTCTGAGCCAAACTTATGATTGTGATTTGAAGAATG	925
Db	2115	GCGAATCGCACTGAAAACGCATCGATGCAACGGTGTCGTGTAACGCCAACAAATG	2171
RESULT 15	A63783	4435 bp	PAT 12-MAR-1998
LOCUS	Sequence 4 from Patent WO9723624.	DNA	linear
DEFINITION	A63783		
ACCESSION	A63783		
VERSION	A63783.1	GI:3717355	
KEYWORDS	unidentified.		
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 4435)		
AUTHORS	Herman-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M., Sumar,N. and Ford,J.		
TITLE	NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY		
JOURNAL	Patent: WO 9723624-A 4 03-JUL-1997;		
COMMENT	ST GEORGE S HOSPITAL MEDICAL S (GB)		
FEATURES	Other publication AU 1202797 19970717. Location/Qualifiers		
source	1..4435		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	904 a 1289 c 1288 g 954 t		
ORIGIN			
Query Match 31.7%; Score 297; DB 6; Length 4435;			
Best Local Similarity 58.2%; Pred.No. 2e-76;			
Matches 522; Conservative 0; Mismatches 375; Indels 0; Gaps 0;			
Qy	29	TCGCGGTCATCGTGGTTGGTTGGATCTGCATTGTCCGCAAGCTTCAGGAAACAAGGTT	88
Db	1275	TCGCGGTCATCGGGGCTGGTCCGCTCAGCGCTCGTAGATTGAGGCCGCGGGT	1334
Qy	89	TCACCAATCTGTTCTTAAACAACA CCGCGAGCTTGATCTCACTCGTCAAAGCCGATGTTG	148
Db	1335	TCACCAATCTCATTTGTCGATCACGCCGATGAGATTGATCTCAGCGGACCGAGCCGCAACGT	1394
Qy	149	AATCCTTTCTTCTCAGAGAAGCCAGTTTATGTATCCTAGCAGCAGCTAAAGTTGGTG	208
Db	1395	TTGATTTTGTCTGAGACAAGCAACAGGTGATCATCGATGCGCGCGCAGCGGTGCGCG	1454
Qy	209	GTAATTCACGCTAACCAACACCTATCCTGCTGANTTCATTTGTTGTTGTCATCTCCAGATTTCAGA	268
Db	1455	GCATCATGCGGAATAACACCTATCCCGCGAGCTTCTTGTCGGAANAACCTCCGAATCCAGA	1514
Qy	269	CCAATGTGATCCACTCTGATATGAGACACGGTGTGAAGAAGCTTCTCTTCCTTGGATCAT	328
Db	1515	CCAATTTGCTCGACGCACTGTGCGCGTGTGTCGCGCGGCTCCTTTTCTCGGTTCTG	1574
Qy	329	CCTGCATTTACCCTAAATTTTCTCCTCAGCCAAATTCCTGAGTCTGCTTTCTGTTAAACAGCAT	388
Db	1575	CATGCATCTACCCGAGTAGCTCTCCGCAACTATCCACGAGAGTGCTTTATTCACCTTGGCC	1534
Qy	389	CGCTTGAAACCAACTAATAGTGGTATGCTATTGCTAAGATCGCTGGGAATTAAGACTTGTC	448
Db	1635	CTTTGGAGCCCAACCAACGACCGGTATGCGATCGCAAGATCGCGGTATCCTCGCAAGTTC	1694
Qy	449	AGGCTTATAGGATTCAGCAGGATGGATGCAATCTCTGCGATGCCCTACTAATCTCTATG	508
Db	1695	AGGCGGTATGGCGCAATATGGGCTGGCGTGGATCTCTGCGATGCCGACTTAACCTCTACG	1754
Qy	509	GTCTTAATGACAATTTCCACCCGAGAAATTTCTCATGTGCTTCTCTGCTCTTATCAGGAGGT	568
Db	1755	GACCCGCGACAACCTTCTCCCGTCCGGTTCGATCTCTTGC CGCGCTCATCCGTCGAT	1814
Qy	569	TCACAGAGGCCAAGTAGATGZAGCGGAGGAATTTGTGTGTGGGTACAGGTAGTCCGT	628



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:04:31 ; Search time 531 Seconds  
(without alignments)  
59.476 Million cell updates/sec

Title: US-10-089-014-1  
Perfect score: 1646  
Sequence: 1 MSDKSAKIFVAGHRLVGS.....SLRDGLSTQDYWLKNVCR 312

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues

Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW PUB.pap.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT NEW PUB.pap.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW PUB.pap.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW PUB.pap.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW PUB.pap.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US03\_PUBCOMB.pap.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW PUB.pap.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW PUB.pap.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1515	92.0	314	10	US-09-318-271-2
2	1241	75.4	310	10	US-09-734-569-148
3	937.5	57.0	321	10	US-09-318-271-4
4	671	40.8	162	10	US-09-734-569-16
5	463.5	28.2	321	10	US-09-318-271-6
6	249	15.1	311	9	US-09-738-626-3862
7	236	14.3	374	9	US-10-168-066-7
8	215	13.1	420	9	US-10-028-072-260
9	215	13.1	420	9	US-10-121-049-260
10	215	13.1	420	9	US-10-123-904-260
11	215	13.1	420	9	US-10-140-470-260
12	215	13.1	420	9	US-10-175-746-260
13	215	13.1	420	9	US-10-176-918-260
14	215	13.1	420	9	US-10-176-921-260
15	215	13.1	420	9	US-10-227-884-240
16	215	13.1	420	9	US-10-137-865-260
17	215	13.1	420	9	US-10-140-474-260
18	215	13.1	420	9	US-10-142-431-260
19	215	13.1	420	9	US-10-143-114-260

20	215	13.1	420	9	US-10	Sequence 240, App
21	215	13.1	420	9	US-10	Sequence 260, App
22	215	13.1	420	9	US-10-142-419-260	Sequence 260, App
23	215	13.1	420	9	US-10-218-631-240	Sequence 240, App
24	215	13.1	420	9	US-10-230-338-240	Sequence 260, App
25	215	13.1	420	9	US-10-123-262-260	Sequence 260, App
26	215	13.1	420	9	US-10-142-423-260	Sequence 260, App
27	215	13.1	420	9	US-10-230-414-240	Sequence 260, App
28	215	13.1	420	9	US-10-121-050-260	Sequence 260, App
29	215	13.1	420	9	US-10-141-755-260	Sequence 260, App
30	215	13.1	420	9	US-10-143-032-260	Sequence 260, App
31	215	13.1	420	9	US-10-123-108-260	Sequence 260, App
32	215	13.1	420	9	US-10-123-236-260	Sequence 260, App
33	215	13.1	420	9	US-10-123-261-260	Sequence 260, App
34	215	13.1	420	9	US-10-140-921-260	Sequence 260, App
35	215	13.1	420	9	US-10-140-928-260	Sequence 260, App
36	215	13.1	420	9	US-10-216-159A-240	Sequence 260, App
37	215	13.1	420	9	US-10-121-045-260	Sequence 260, App
38	215	13.1	420	9	US-10-123-292-260	Sequence 260, App
39	215	13.1	420	9	US-10-123-903-260	Sequence 260, App
40	215	13.1	420	9	US-10-124-819-260	Sequence 260, App
41	215	13.1	420	9	US-10-124-822-260	Sequence 260, App
42	215	13.1	420	9	US-10-140-925-260	Sequence 260, App
43	215	13.1	420	9	US-10-160-438-260	Sequence 260, App
44	215	13.1	420	9	US-10-218-849-240	Sequence 240, App
45	215	13.1	420	9	US-10-227-873-240	Sequence 240, App

## ALIGNMENTS

RESULT 1  
US-09-318-271-2  
; Sequence 2, Application US/09318271A  
; Patent No. US20020012979A1  
; GENERAL INFORMATION:  
; APPLICANT: Berry, Alan  
; APPLICANT: Running, Jeffrey A.  
; APPLICANT: Severnson, David K.  
; APPLICANT: Burlingame, Richard P.  
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"  
; FILE REFERENCE: 3161-24  
; CURRENT APPLICATION NUMBER: US/09/318,271A  
; EARLIER FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 60/125,073  
; EARLIER FILING DATE: 1999-03-17  
; EARLIER APPLICATION NUMBER: 60/125,054  
; EARLIER FILING DATE: 1999-03-18  
; EARLIER APPLICATION NUMBER: 60/088,549  
; EARLIER FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 314  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-318-271-2

Query Match	92.0%	Score 1515;	DB 10;	Length 314;
Best Local Similarity	92.5%	Pred. No. 5.9e-142;		
Matches 298;	Conservative	1; Mismatches	5;	Indels 18; Gaps 5;
QY	1	MSDKSAKIFVAGHRLVGS	1	MSDKSAKIFVAGHRLVGS
DB	1	MSDKSAKIFVAGHRLVGS	1	MSDKSAKIFVAGHRLVGS
QY	61	VILAAKVGIIHANNYTPADFIGVNLQIQTNIHSAYEHGVKKLFLGSSCIYPKAPOP	120	VILAAKVGIIHANNYTPADFIGVNLQIQTNIHSAYEHGVKKLFLGSSCIYPKAPOP
DB	61	VILAAKVGIIHANNYTPADFIGVNLQIQTNIHSAYEHGVKKLFLGSSCIYPKAPOP	120	VILAAKVGIIHANNYTPADFIGVNLQIQTNIHSAYEHGVKKLFLGSSCIYPKAPOP
QY	121	IPESALLTASLPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPHS	180	IPESALLTASLPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPHS
DB	121	IPESALLTASLPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPHS	180	IPESALLTASLPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPHS



QY 181 HVLPAALRRPHEAKVGAEEVVGWGTGS-----PLR-EFLHVDLADACVFLDRYS 231  
 Db 180 HVLPAALRRPHEAKV-----WSGSCGVGVKVPLEKFLHVDLADACVFLDRIQ 232  
 QY 232 -GLEHNVIGSQEVTIRELAELVKEVVFEGKLGWCTKPDGTGPRKLMSSKLASLGWTF 290  
 Db 233 RGLHNVIGSQEVTIRELAELVKEVVFEGKLGWCTKPDGTGPRKLMSSKLASLGWTF 292  
 QY 291 KVSRLDGLSQTWDYLNKVCNR 312  
 Db 293 KVSRLDGLSQTWDYLNKVCNR 314

RESULT 2  
 US-09-734-569-148  
 ; Sequence 148, Application US/09734569  
 ; Patent No. US20020064816A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lerchl, Jens  
 ; APPLICANT: Renz, Andreas  
 ; APPLICANT: Ehrhardt, Thomas  
 ; APPLICANT: Reindl, Andreas  
 ; APPLICANT: Cirpus, Petra  
 ; APPLICANT: Bischoff, Friedrich  
 ; APPLICANT: Frank, Markus  
 ; APPLICANT: Freund, Annette  
 ; APPLICANT: Duwenig, Elke  
 ; APPLICANT: Schmidt, Ralf-Michael  
 ; APPLICANT: Reski, Ralf  
 ; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved  
 ; FILE REFERENCE: BASF-NAB-1332-99-US  
 ; CURRENT APPLICATION NUMBER: US/09734,569  
 ; PRIOR FILING DATE: 2001-05-24  
 ; PRIOR APPLICATION NUMBER: US 60/171,101  
 ; NUMBER OF SEQ ID NOS: 181  
 ; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1  
 ; SEQ ID NO 148  
 ; LENGTH: 310  
 ; TYPE: PRT  
 ; ORGANISM: Physcomitrella patens  
 US-09-734-569-148

Query Match 75.4%; Score 1241; DB 10; Length 310;  
 Best Local Similarity 75.5%; Pred. No. 8.4e-115;  
 Matches 231; Conservative 37; Mismatches 36; Indels 2; Gaps 1;  
 QY 3 DSAKIFVAGHGLVGSIAVRKLEQOGFTNLVLTAEHDLTRQADVESFPOEKPVYVI 62  
 Db 4 DKAIFVAGHGLVGSIAVRKLEQOGFTNLVLTAEHDLTRQADVESFPOEKPVYVI 61  
 QY 63 LAAAKVGIIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIP 122  
 Db 62 LAAAKVGIIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIP 121  
 QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLGPNNDHFNHSHV 182  
 Db 122 EESLLTGLPEATNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLGPNNDHFNHSHV 181  
 QY 183 LPALMRPFHEAKVGAEEVVGWGTGSPLRFLHVDLADACVFLDRYSGLHNVIGSQ 242  
 Db 182 LPALMRPFHEAKVGAEEVVGWGTGSPLRFLHVDLADACVFLDRYSGLHNVIGSQ 241  
 QY 243 EVTIRELAELVKEVVFEGKLGWCTKPDGTGPRKLMSSKLASLGWTFKVSRLDGLSQT 302  
 Db 242 EVSIRELAELVKEVVFEGKLGWCTKPDGTGPRKLMSSKLASLGWTFKVSRLDGLSQT 301  
 QY 303 DWYLNK 308  
 Db 302 KWCYCN 307

RESULT 3  
 US-09-318-271-4  
 ; Sequence 4, Application US/09318271A  
 ; Patent No. US20020012979A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bert, Alan  
 ; APPLICANT: Running, Jeffrey A.  
 ; APPLICANT: Severson, David K.  
 ; APPLICANT: Burlingame, Richard P.  
 ; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"  
 ; FILE REFERENCE: 3161-24  
 ; CURRENT APPLICATION NUMBER: US/09/318,271A  
 ; CURRENT FILING DATE: 1999-05-25  
 ; EARLIER APPLICATION NUMBER: 60/125,073  
 ; EARLIER FILING DATE: 1999-03-17  
 ; EARLIER APPLICATION NUMBER: 60/125,054  
 ; EARLIER FILING DATE: 1999-03-18  
 ; EARLIER APPLICATION NUMBER: 60/088,549  
 ; EARLIER FILING DATE: 1998-06-08  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; TYPE: PRT  
 ; LENGTH: 321  
 ; ORGANISM: Escherichia coli  
 US-09-318-271-4

Query Match 57.0%; Score 937.5; DB 10; Length 321;  
 Best Local Similarity 56.9%; Pred. No. 1.1e-84;  
 Matches 177; Conservative 58; Mismatches 67; Indels 9; Gaps 1;  
 QY 7 KIFVAGHGLVGSIAVRKLEQOGFTNLVLTAEHDLTRQADVESFPOEKPVYVILAAA 66  
 Db 5 RVFIAGHGMVGSIAVRKLEQOGFTNLVLTAEHDLTRQADVESFPOEKPVYVILAAA 64  
 QY 67 KVGGIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIPESAL 126  
 Db 65 KVGGIHANNTYPADFIGVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIPESAL 124  
 QY 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLGPNNDHFNHSHVLPAL 186  
 Db 125 LGTLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLGPNNDHFNHSHVLPAL 184  
 QY 187 MRFHEAKVGAEEVVGWGTGSPLRFLHVDLADAC-----VFLDRYSGLHNV 237  
 Db 185 LRRFHEATQANAPDVVVGWGTGSPLRFLHVDLADAC-----VFLDRYSGLHNV 244  
 QY 238 IGSQGVITIRELAELVKEVVFEGKLGWCTKPDGTGPRKLMSSKLASLGWTFKVSRLD 297  
 Db 245 VGTGVDTTIRDAVQATKAVGVYGRVVFDAKSPDGTGPRKLLDVTRLHQLGWYHISLEAG 304  
 QY 298 LSQTYDWYLNK 308  
 Db 305 LASTYQWFLN 315

RESULT 4  
 US-09-734-569-16  
 ; Sequence 16, Application US/09734569  
 ; Patent No. US20020064816A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lerchl, Jens  
 ; APPLICANT: Renz, Andreas  
 ; APPLICANT: Ehrhardt, Thomas  
 ; APPLICANT: Reindl, Andreas  
 ; APPLICANT: Cirpus, Petra  
 ; APPLICANT: Bischoff, Friedrich  
 ; APPLICANT: Frank, Markus  
 ; APPLICANT: Freund, Annette  
 ; APPLICANT: Duwenig, Elke  
 ; APPLICANT: Schmidt, Ralf-Michael  
 ; APPLICANT: Reski, Ralf

5 17:~



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2003, 03:32:11 ; Search time 191 Seconds  
(without alignments)  
6607.695 Million cell updates/sec

Title: US-10-089-014-2  
Perfect score: 936  
Sequence: 1 atgtctgacaaatctgccaa.....tgaagaatgtttgcaaccga 936

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 845702 seqs, 674182571 residues  
Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	857	91.6	1583	10	US-09-318-271-1
2	391	41.8	1490	10	US-09-734-569-147
3	275.2	29.4	966	9	US-09-894-844-49
C 4	269.8	28.8	14187	9	US-10-114-170-121
C 5	265	28.3	536165	9	US-09-939-964-1
6	262.2	28.0	966	10	US-09-318-271-3
7	213.6	22.8	701	10	US-09-734-569-15
8	145.8	15.6	933	10	US-09-962-805-2
9	101.6	10.9	1340	10	US-09-318-271-5
C 10	59.8	6.4	282	10	US-09-294-0938-2058
C 11	58.8	6.3	447	9	US-10-025-380-296
C 12	58.8	6.3	447	10	US-09-922-217-296
C 13	58.8	6.3	447	10	US-09-833-263-296
C 14	44.8	4.8	656	10	US-09-815-343-253
C 15	43.8	4.7	846	10	US-09-770-445-638
16	38.6	4.1	492	10	US-09-920-300A-1521
17	38.6	4.1	492	12	US-10-033-528-1521
18	38.4	4.1	1332	9	US-09-938-842A-190
19	38	4.1	413	9	US-10-184-644-156

Sequence 156, Appli	413	9	US-10-184-634-156
Sequence 1, Appli	925	9	US-09-735-056-1
Sequence 26997, A	494	9	US-09-918-995-26997
Sequence 154, Appl	4495	9	US-10-097-340-154
Sequence 98, Appli	4495	9	US-10-171-311-98
Sequence 96, Appli	4637	9	US-10-171-311-96
Sequence 7, Appli	5503	12	US-10-044-090-7
Sequence 60, Appli	1512	9	US-10-001-857-60
Sequence 8082, Ap	482	9	US-09-796-632-8082
Sequence 8082, Ap	34.8	3.7	US-10-040-862-8082
Sequence 346, App	671	9	US-10-184-644-346
Sequence 346, App	671	9	US-10-184-634-346
Sequence 6231, Ap	1968	10	US-09-815-242-6231
Sequence 281, App	1990	9	US-09-809-391-281
Sequence 259, App	2024	9	US-10-028-072-259
Sequence 259, App	2024	9	US-10-121-049-259
Sequence 259, App	2024	9	US-10-123-904-259
Sequence 259, App	2024	9	US-10-140-470-259
Sequence 259, App	2024	9	US-10-175-746-259
Sequence 259, App	2024	9	US-10-176-918-259
Sequence 259, App	2024	9	US-10-227-884-239
Sequence 239, App	2024	9	US-10-176-921-239
Sequence 259, App	2024	9	US-10-137-865-259
Sequence 259, App	2024	9	US-10-140-474-259
Sequence 259, App	2024	9	US-10-142-431-259
Sequence 259, App	2024	9	US-10-143-114-259

ALIGNMENTS

RESULT 1  
US-09-318-271-1  
; Sequence 1, Application US/09318271A  
; Patent No. US20020012979A1  
; GENERAL INFORMATION:  
; APPLICANT: Berry, Alan  
; APPLICANT: Running, Jeffrey A.  
; APPLICANT: Severson, David K.  
; APPLICANT: Burlingame, Richard P.  
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"  
; FILE REFERENCE: 3161-24  
; CURRENT APPLICATION NUMBER: US/09/318,271A  
; EARLIER FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 60/125,073  
; EARLIER FILING DATE: 1999-03-17  
; EARLIER APPLICATION NUMBER: 60/125,054  
; EARLIER FILING DATE: 1999-03-18  
; EARLIER APPLICATION NUMBER: 60/088,549  
; EARLIER FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1583  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (49)..(990)  
US-09-318-271-1

Query Match	91.6%	Score 857;	DB 10;	Length 1583;
Best Local Similarity	97.1%	Pred. No. 2.8e-274;		
Matches 918;	Conservative	0; Mismatches 15;	Indels 12;	Gaps 4;
Qy	1	ATGTCTGACAAATCTGCCAAATCTTCGTCGGGTCATCGTGGTGTGGTGGATCTGCC	60	
Db	49	ATGTCTGACAAATCTGCCAAATCTTCGTCGGGTCATCGTGGTGTGGTGGATCTGCC	108	
Qy	61	ATTGTCCGCAAGTTCAGGAACCAAGGTTTACCAATCTCGTTCTTAAACACACGCCGAG	120	
Db	109	ATTGTCCGCAAGTTCAGGAACCAAGGTTTACCAATCTCGTTCTTAAACACACGCCGAG	168	





